AFOSR-TR- 80-1293

LEVEL

ADMISSIBILITY OF ESTIMATORS IN THE ONE PARAMETER EXPONENTIAL FAMILY AND IN MULTIVARIATE LOCATION PROBLEMS

by

Dan A. Ralescu

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
for the degree Doctor of Philosophy
in the Department of Mathematics
Indiana University

September, 1980

80 12 22 194

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR-TR-80-1295	2. GOVT ACCESSION NO. AD-A093 188	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ADMISSIBILITY OF ESTIMATORS IN THE ONE PARAMETER EXPONENTIAL FAMILY AND IN MULTIVARIATE LOCATION PROBLEMS.		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dan A. Ralescu		8. CONTRACT OR GRANT NUMBER(s) AFOSR-76-2927
9. PERFORMING ORGANIZATION NAME AND ADDRESS Indiana University Department of Mathematics Bloomington, Indiana 47405		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (AFSC) Bolling Air Force Base, D.C. 20332		12. REPORT DATE ~~September~~ Sep 1980
		13. NUMBER OF PAGES 70
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Admissibility, minimaxity, exponential family, affine estimators, truncated probability density, invariant estimators, truncated parameter space.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let X be a random variable with probability density function $f(x) = \beta(\theta)e^{\theta x}$ with respect to some σ-finite measure $\mu; \theta \in \Theta = \{\theta : \beta^{-1}(\theta) = \int e^{\theta x} \, d\mu(x) < \infty\}$. Sufficient conditions are obtained for the admissibility of nonlinear estimators of the form $\delta(X) = (aX + b)/(cX + d)$ for the problem of estimating an arbitrary piecewise continuous, locally integrable function $g(\theta)$

DD FORM 1473 1 JAN 73 (over)

UNCLASSIFIED

20. (continued)

with squared error loss. Several applications of the main result are given and some new nonlinear admissible estimators are discovered. The problem is also studied for the case when the parameter space is truncated, that is, when $\theta \in \Theta_o = \{\theta : \theta \le \theta_o\} \subset \Theta$, θ_o being an interior point of Θ. The minimaxity of linear estimators of the form $aX + b$ in estimating an arbitrary differentiable function $g(\theta)$ is also investigated.

In the multivariate case the classes of estimators which improve upon the best invariant estimator are considered. Let X be a p-dimensional random vector having a density of the form $f(x - \theta)$ where $\theta \in \mathbb{R}^p$ is a location parameter. For $p \ge 3$, different classes of estimators $\delta(X)$ which are uniformly better than the best invariant estimator $\delta_o(X) = X$ are obtained when the loss function is of the type

$$L(\theta, \delta(x)) = \sum_{i=1}^{p} c_i \, (\theta_i - \delta_i(x))^2 \text{ , where } c_1, \ldots, c_p \text{ are}$$

given positive constants. It is shown that $\delta = \delta_1 * \xi$, where δ_1 is an estimator which improves upon δ_o outside of a compact set, ξ is a suitable probability density in $\mathbb{R}^p$, and $*$ denotes the convolution. Some examples of densities ξ (such as truncated densities) which generate estimators which improve upon δ_o are given, and some problems of further research interest are also stated.

Accepted by the faculty of the Graduate School, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Madan L. Puri, Chairman

Grahame Bennett

Victor W. Goodman

(for Daniel Maki)

Daniel P. Maki

A

ACKNOWLEDGEMENTS

It is a pleasure to express my gratitude to Professor Madan Puri for his invaluable advice during the past years. Without his constant encouragements and interest in my work this thesis could not have been written.

I am indebted to Professor Daniel Maki for the effort and encouragements he provided.

My special thanks are due to Professors Grahame Bennett and Victor Goodman for serving in my thesis committee.

Finally, I would like to thank my wife Anca for her understanding and moral support during all these years.

This work was supported in part by the Air Force Office of Scientific Research, AFSC USAF, Grant No. AFOSR-76-2927.

Admissibility of estimators in the one parameter exponential family and in multivariate location problems

by

Dan A. Ralescu

ABSTRACT

We are concerned with problems related to admissibility and minimaxity of estimators in the one parameter exponential family, and with classes of estimators which improve upon the best invariant estimator in multivariate location problems (dimensions $p \geq 3$). > or =

In connection with admissibility problems we give sufficient conditions for the admissibility of nonlinear estimators theta $(aX + b)/(cX + d)$ in estimating an arbitrary function $g(\theta)$ with quadratic loss. More precisely, let X be a random variable with probability density function $f_\theta(x) = \beta(\theta)e^{\theta x}$ with respect to some σ-finite measure μ ; $\theta \in \Theta = \{\theta : \beta^{-1}(\theta) = \int e^{\theta x} d\mu(x) < \infty\}$. Sufficient conditions are obtained for the admissibility of nonlinear estimators of the form $\delta(X) = (aX + b)/(cX + d)$ for the problem of estimating an arbitrary piecewise continuous, locally integrable function $g(\theta)$ with squared error loss.

Several applications of the main result are given and some new nonlinear admissible estimators are discovered. In particular, if $X_1, X_2, \ldots, X_n$ is a sample from the exponential density

$\lambda e^{-\lambda x} I_{(0,\infty)}(x)$, $\lambda > 0$, we show that $(n - 2)/(X + k)$ is admissible in estimating λ , for every $k \geq 0$, where

$X = \sum_{i=1}^{n} X_i$.

This problem is also studied for the case when the parameter space is truncated, that is, when $\theta \in \Theta_o = \{\theta : \theta \leq \theta_o\} \subset \Theta$, θ_o being an interior point of Θ .

Problems related to minimaxity of linear estimators are also investigated. If X has a density belonging to the exponential family, we give sufficient conditions for the minimaxity of estimators of the form $aX + b$ in estimating arbitrary differentiable function $g(\theta)$.

In the multivariate case we concentrate on classes of estimators which improve upon the best invariant estimator. Let X be a p-dimensional random vector having a density of the form $f(x - \theta)$ where $\theta \in \mathbb{R}^p$ is a location parameter. For $p \geq 3$, different classes of estimators $\delta(X)$ which are uniformly better than the best invariant estimator $\delta_o(X) = X$ are obtained when the loss function is of the type

$L(\theta,\delta(x)) = \sum_{i=1}^{p} c_i \, (\theta_i - \delta_i(x))^2$, where $c_1,\ldots,c_p$ are given positive constants.

It is shown that $\delta = \delta_1 * \xi$, where δ_1 is an estimator which improves upon δ_o outside of a compact set, ξ is a suitable probability density in $\mathbb{R}^p$, and $*$ denotes the convolution. We give some examples of densities ξ (such as truncated densities) which generate estimators which improve upon δ_o , and we also state some problems which could be of further research interest.

Madan L. Puri, Chairman

Victor W. Goodman

Grahame Bennett

Joseph Stampfli for Daniel Maki

Daniel P. Maki

TABLE OF CONTENTS

CHAPTER 0

PRELIMINARIES

In this chapter we describe the general decision problem, the concepts of admissibility and minimaxity, some related background and the problems investigated in Part I of the thesis.

Let X be a random variable with distribution function

$$(0.1) \qquad F_\theta(x) = \int_{-\infty}^{x} f_\theta(t)\,d\mu(t)$$

depending on an unknown parameter $\theta \in \Theta \subseteq R$. The set Θ which is assumed known is the parameter space. The quantity $f_\theta(x)$ is the density of $F_\theta(x)$ with respect to a σ-finite measure μ . An important problem in statistics is to estimate θ or a function $g(\theta)$ of θ on the basis of an observation X (or series of observations) on F_θ . This is done by determining a rule which for each set of values of observations, specifies what decision should be taken. Mathematically, such a rule is a function δ , which to each possible value x of X assigns a decision $d = \delta(x)$, that is, a function whose domain is a set of values of X and whose range is a set of possible decisions.

To see how δ should be chosen one has to compare the consequences of using different rules. Suppose that the consequence of taking a decision d in estimating $g(\theta)$ is a loss which can be expressed as a non-negative real number $L(g(\theta), d)$. Then the long term average loss that would result from the use of δ in a number of repetitions of the experiment is the expectation $E_\theta[L(g(\theta),\delta(X))]$ evaluated under the assumption that f_θ is the true density of X . This expectation which depends on the decision rule δ and the density f_θ is called the risk function of δ , and we shall denote it by $R(g(\theta),\delta)$. Thus

$$(0.2) \qquad \begin{aligned} R(g(\theta),\delta) &= E_\theta[L(g(\theta),\delta(X))] \\ &= \int L(g(\theta),\delta(x))f_\theta(x)d\mu(x) \ . \end{aligned}$$

By basing the decision on the observations, the original problem of choosing a decision d with the loss function $L(g(\theta),d)$ is thus replaced by choosing δ with the average loss $R(g(\theta),\delta)$.

Ideally, one would like to select δ which minimizes the risk function (0.2) for all $\theta \in \Theta$. Unfortunately, in general such "best" decision rules do not exist. So sometimes one is led to considering restricted classes of decision procedures which possess a certain degree of impartiality such as unbiasedness, invariance, minimax, Bayes', etc. in the hope that within such a restricted class there may exist a procedure which is uniformly best. However, there are situations in which there exists a decision procedure δ_0 with uniformly minimum risk among all invariant or unbiased procedures, but where there exists a procedure δ , not possessing this impartiality property and preferable to δ_0 (see, for example, Lehmann (1959), pages 24 and 26, problems 14 and 16). Thus the approaches based on the principles of unbiasedness or invariance could be unreliable, and for different reasons, the approaches based on the minimax or Bayes principle could also be far from satisfactory. Thus, as a first step, one considers the possibility of not insisting on a unique solution but asking only how far a decision problem can be reduced without loss of relevant information. This leads to the concept of admissibility.

<u>Definition 1</u>. A decision procedure δ_0 is said to be <u>inadmissible</u> if there exists another procedure δ_1 which dominates it in the sense that

$$
(0.3)\qquad \begin{aligned} &R(g(\theta),\delta_1) \le R(g(\theta),\delta_0) \quad \text{for all } \theta \in \Theta \\ &R(g(\theta),\delta_1) < R(g(\theta),\delta_0) \quad \text{for at least one } \theta \in \Theta . \end{aligned}
$$

δ_0 is called <u>admissible</u> if no such dominating δ_1 exists. Thus a decision procedure δ_0 can be eliminated from consideration if there exists a procedure δ_1 which dominates it.

A class T of decision procedures is said to be complete if for any δ_0 not in T , there exists a δ_1 in T which dominates it.

The importance of admissible procedures lies in the fact that under suitable assumptions on the loss function and the density function, the admissible procedures form a complete class.

In fact, if a minimal complete class exists, it consists exactly of the totality of the admissible procedures (and consequently there is no need to look outside this class to find an estimation procedure, for one can just do as well inside the class). However, the general question of resolving the admissibility of all estimates measured with respect to a suitable loss function (say, a quadratic one - frequently used in practice) is intrinsically difficult. One therefore concentrates on the investigation of whether some of the commonly employed estimates are admissible.

One of the earliest papers in this direction is due to Hodges and Lehmann (1951) who used the Cramér-Rao inequality for the variance of an estimator of the parameter θ to obtain a criterion which implies the admissibility of point estimators when the loss is proportional to the square of the error of the estimate. Their method which involves the solution of a differential inequality, is applied to certain problems involving the binomial, Poisson, normal, and chi square distributions, and the unique admissible minimax estimator is obtained in each case. Simultaneously Girshick and Savage (1951), while investigating related problems in relatively greater generality, proved (among other results) that if the distribution of X belongs to a one parameter exponential family where $f_\theta(x) = \beta(\theta)e^{\theta x}$, and if the loss function is the same as in Hodges and Lehmann (1951), then X is an admissible (minimax) estimator of $E_\theta X$ provided $-\infty < \theta < \infty$. Later Karlin (1958) proved an interesting and surprising result - that for the exponential family given above, aX for any a satisfying $0 < a \leq 1$ is an admissible estimator of $E_\theta X$ whenever μ possesses positive measure in the regions $x \geq 0$ and $x \leq 0$ and $\Theta = (-\infty,\infty)$. On the other hand, for any $a > 1$, aX is inadmissible. In view of the fact that any contraction of X $(aX,\ 0 < a \leq 1)$ is admissible, it seems surprising that in practice one always uses the extreme estimate of this kind. The criterion of unbiasedness traditionally has dominated the choice of an estimator If the parameter space Θ of θ is not the full infinite interval, then the problem of admissibility of aX becomes quite complicated. It becomes more so if θ ranges over a finite interval. In this case the analysis seems to depend on the rate at which $\beta(\theta)$ tends to 0 as θ approaches its boundary. (For details see Karlin (1958)).

Later Ping (1964) and Gupta (1966) gave sufficient conditions for the admissibility of the estimators of the form $aX + b$ for the problem considered in Karlin cited above. More recently, Ghosh and Meeden (1977) considered the problem of estimating a piece wise continuous function $\gamma(\theta)$, not necessary the mean, by an estimator of the form $a X + b$, and provided sufficient conditions for the admissibility of such estimators. All these papers deal with linear or affine estimators. However, there are important problems where the class of estimators studied include nonlinear estimators. In particular, this is the case in estimating a function of the scale parameter in a Gamma density, or a function of the variance in a normal density. The admissibility of particular nonlinear estimators of the form c/X , where c is a constant, has been studied by Ghosh and Singh (1970), in estimating the parameter of an exponential density. However, as of now, there are no general results available dealing with non-linear estimators which would apply to a broader class of densities.

In Chapter 1, Section 1.2 we give sufficient conditions for the admissibility of nonlinear estimators of the form $\delta(X) = (aX + b)/(cX + d)$, in estimating an arbitrary function $g(\theta)$, with squared error loss. The results obtained include the results of Karlin (1958), Ghosh and Singh (1970), and Ghosh and Meeden (1977), among others.

As a Corollary, we give sufficient conditions for the admissibility of estimators of the form $\delta(X) = c/X$.

In Section 1.3 we give several examples of nonlinear admissible estimators, especially in estimating a function of the scale parameter in a Gamma or normal density. Some new admissible estimators are discovered. A surprising example shows the "almost inadmissibility" (in a sense to be made precise later) of the commonly used estimator of the variance in a normal $(0,\sigma^2)$ density.

In Section 1.4 we extend the results of Katz (1961) and of Ghosh and Meeden (1977) concerning admissibility when the parameter space is truncated. We derive admissible estimators of the form $(aX + b)/(cX + d) + \Phi(X)$, where $\Phi(X)$ is a "correction" due to the truncation.

The problem of admissibility is related to the problem of finding minimax estimators. To define the latter concept, let $R(g(\theta),\delta)$ denote, as before, the risk associated with the estimator $\delta(X)$.

An estimator δ_o is called minimax, if:

$$\sup_{\theta\in\Theta} R(g(\theta),\delta_o) = \inf \sup_{\theta\in\Theta} R(g(\theta),\delta) \tag{0.4}$$

where the infimum is taken over all estimators $\delta(X)$ of $g(\theta)$. Intuitively, the minimax approach is to choose an estimator which protects against the largest possible risk, when θ varies over Θ. There is a considerable amount of published results on the existence of minimax estimators; see especially Chapter 2 of Wald (1950).

In the case of the one parameter exponential family, Ping (1964) gave sufficient conditions for the minimaxity of affine estimators of the form $\delta(X) = aX + b$, in estimating the mean $g(\theta) = E_\theta X$, under the normalized squared error loss

$$L(g(\theta),\delta(x)) = \frac{(g(\theta) - \delta(x))^2}{\operatorname{var}_\theta X} \tag{0.5}$$

where $\operatorname{var}_\theta X$ is the variance of X.

In Chapter 2, Section 2.1, we generalize this result and we give sufficient conditions for the minimaxity of $aX + b$ in estimating an arbitrary (differentiable) function $g(\theta)$ when the loss function is (0.5).

In Section 2.2 we give some new examples of minimax estimators, in estimating a function $g(\theta)$ different from the mean, $E_\theta X$. The presence of affine minimax estimators arises especially in estimating a function of the scale parameter in a Gamma density.

In Chapters 1 and 2, the observations are considered to be random variables, and the parameter to be estimated is one-dimensional.

In multivariate estimation problems, the situation changes considerably. The standard example is when the observation $X = (X_1, X_2, \ldots, X_p)$ has a p-variate normal distribution with mean θ and covariance matrix I_p (the $p \times p$ identity matrix).

It was a surprising result when Stein (1955) and, later, James and Stein (1960) showed that, in dimensions $p \geq 3$, the best invariant estimator $\delta(X) = X$ of a multivatiate normal mean, is <u>inadmissible</u>. They found an estimator which strictly dominates δ. More precisely, they proved that δ is inadmissible if and only if $p \geq 3$.

Since then, much work has been done in the direction of proving inadmissibility of the best invariant estimator, for estimation problems in a relatively general framework. Based on results of Farrell (1964), many contributions to this subject were made by Brown ((1966),(1975)). In Brown (1978) a heuristic approach is given to prove admissibility and inadmissibility of estimators in a wide variety of multivariate problems.

The work of Stein and Brown suggest a new problem, namely that of finding estimators which are better than the best invariant estimator when sampling from a location parameter family.

In James and Stein (1960) estimators for the mean of a multivariate normal distribution are given, which are better than the best invariant estimator X, where X is an observation of the distribution.

Baranchik (1964) found a larger class of estimators, better than X, which include the James-Stein estimators.

Until 1974, estimators which were better than the best invariant estimator were only available for the mean vector of a multivariate normal distribution. Then, Strawderman (1974) and Berger (1975) found minimax estimators which are better than the best invariant estimators, when sampling from certain spherically symmetric unimodal distributions. Later, more results along these lines were obtained, by Brandwein and Strawderman (1978), Brandwein (1979), and Brandwein and Strawderman (1980). Mainly, these results concentrated on describing classes of minimax estimators for the mean of a spherically symmetric distribution, for various classes of loss functions, more general than the quadratic loss.

However, estimators which improve upon the best invariant estimator have been found only in the special cases of normal and spherically symmetric distributions. In Chapter 3 we investigate several types of estimators which improve upon the best invariant estimator when the underlying distributions are not necessarily normal or spherically symmetric and also when the loss function is relatively more general than the quadratic one (frequently considered in literature). We prove that (under suitable assumptions) the convolution of an estimator δ_1, which improves upon δ_o outside of a compact set with a truncated probability density in $\mathbb{R}^p$, gives an estimator δ which improves uniformly upon the best invariant estimator δ_o.

Different examples are given and, according to the convoluting density (such as spherically uniform, truncated densities, and others), different classes of estimators better than δ_o are described.

In Section 3.3, we present some other new classes of estimators which improve upon the best invariant estimator δ_o in higher dimensions. The critical dimensions for which we have improvement depends on the estimator we start with, and which improves upon δ_o outside of a compact set.

Finally, we also state some problems, which were not solved in the present context, but which could be of further research interest.

CHAPTER I

A CLASS OF NONLINEAR ADMISSIBLE ESTIMATORS

In this chapter we investigate the admissibility of non-linear estimators of the form $(aX + b)/(cX + d)$ in the one parameter exponential family $f_\theta(x) = \beta(\theta)e^{\theta x}$, in estimating an arbitrary function $g(\theta)$ with quadratic loss. Particular cases of the estimators of the form c/X are also studied and several examples of nonlinear admissible estimators of the form $(aX + b)/(cX + d)$ and c/X are given. We also consider the problem of admissibility when the parameter space is truncated and derive the admissible estimators of the form $(aX + b)/(cX + d) + \phi(X)$, where $\phi(X)$ is a "correction" due to truncation.

Since our problem originates from Karlin (1958) who studied the admissibility of linear estimators of the form aX in estimating the mean $E_\theta X$ of the one parameter exponential family, and since our results also include those of Ghosh and Meeden (1977), we state the main theorems of Karlin (1958) and, Ghosh and Meeden (1977) in Section 1.1 for appropriate background. In Section 1.2 we present the main theorem dealing with the admissibility of nonlinear estimators of the form $(aX + b)/(cX + d)$ in estimating an arbitrary function $g(\theta)$ with quadratic loss. As a corollary we give sufficient conditions for the admissibility of the estimators of the form c/X. In Section 1.3 we give several examples of nonlinear admissible estimators of the form $(aX + b)/(cX + d)$ and c/X. These examples come especially from estimating a function $g(\lambda)$ in an exponential distribution $\lambda e^{-\lambda x} I_{(0,\infty)}(x)$ and $g(\sigma^2)$ in a normal density $N(0,\sigma^2)$. We also give an example showing the "almost inadmissibility" of the parameter commonly used in estimating the variance in the normal density $N(0,\sigma^2)$. In section 1.4 we derive admissible estimators of the form $(aX + b)/(cX + d) + \phi(X)$ (where $\phi(X)$ is a "correction") in the case when the parameter space is truncated.

1.1. Admissibility of linear estimators

In this section we give for the ease of convenience and appropriate background the main result of Karlin (1958) from where our problem originated. Let the random variable X be distributed according to the probability density

$$(1.1) \qquad dF_\theta(x) = \beta(\theta)e^{\theta x}d\mu(x) ,$$

where μ is a σ-finite measure defined on the real line, θ is an unknown parameter, and we assume that $\theta \in \Theta$, where

$$(1.2) \qquad \Theta = \{\theta \in \mathbb{R} : \int_{-\infty}^{\infty} e^{\theta x}\, d\mu(x) < \infty\} .$$

Since Θ is a convex subset of $\mathbb{R}$, it is an interval of the real line. Let $\bar{\theta}$ and $\underline{\theta}$ be the upper and lower end points of Θ, respectively. Karlin (1958) considered the problem of estimating $g(\theta) = E_\theta X$ from a single observation X and derived sufficient conditions for the admissibility of linear estimators of the form $\delta(X) = aX$, $a \neq 0$ in estimating $g(\theta)$ ($= E_\theta X$) when the loss function is of the form

$$L(g(\theta),\delta(x)) = (g(\theta) - \delta(x))^2 . \tag{1.3}$$

(It may be mentioned that there is no loss of generality in restricting to the case of a single observation because a sufficient statistic for n independent observations from a member of the exponential family (1.1) is the sum of the observations whose distribution is also a member of the exponential family (1.1). (See Blackwell and Girshick (1954, p. 221)). Precisely, Karlin (1958) proved the following theorem.

<u>Theorem 1.1</u>. (Karlin). <u>Let</u> $f_\theta(x) = \beta(\theta)e^{\theta x}$, $\theta \in \Theta$ <u>be the exponential family of densities with respect to a measure</u> μ. <u>If</u>

$$\int_c^b \beta^{-\lambda}(\theta)d\theta \to +\infty \quad \text{as} \quad b \to \bar{\theta} , \tag{1.4}$$

<u>and</u>

$$\int_a^c \beta^{-\lambda}(\theta)d\theta \to +\infty \quad \text{as} \quad a \to \underline{\theta} , \tag{1.5}$$

<u>where</u> c <u>is an interior point of</u> $\Theta = (\underline{\theta},\bar{\theta})$, <u>then</u> $\delta_o(X) = X/(\lambda + 1)$ <u>is an admissible estimator of</u> $g(\theta) = E_\theta X$.

More recently, Ghosh and Meeden (1977) gave sufficient conditions for the admissibility of affine estimators $\delta(X) = aX + b$,

$a \neq 0$ in estimating an arbitrary piecewise continuous, locally integrable function $g(\theta)$ with the quadratic loss (1.3). Specifically, Ghosh and Meeden proved the following theorem.

Theorem 1.2. (Ghosh and Meeden). *Let* $f_\theta(x)$ *be as given in Theorem 1.1. Let*

$$\phi(\theta) = \beta(\theta)\exp\{a^{-1}\int_\alpha^\theta g(t)dt - b\theta/a\} \tag{1.6}$$

where α *is an interior point of* Θ . *If*

$$\int_c^b \phi(\theta)d\theta \to +\infty \quad \text{as} \quad b \to \bar{\theta}\ , \tag{1.7}$$

and

$$\int_a^c \phi(\theta)d\theta \to +\infty \quad \text{as} \quad a \to \underline{\theta}\ , \tag{1.8}$$

where c *is an interior point of* $\Theta = (\underline{\theta}, \bar{\theta})$, *then* $\delta_o(X) = aX + b$ *is an admissible estimator of* $g(\theta)$ (*where* $g(\theta)$ *is an arbitrary, piecewise continuous locally integrable function) when the loss function is given by* (1.3).

These sufficient conditions describe the tail behavior of some improper prior distribution.

Although not studied in the present context, an important problem is related to the converse of Theorem 1.1. More precisely, are these conditions for admissibility also necessary?

In Karlin (1958) it was shown that if one of the integrals in Theorem 1.1 is convergent, then the corresponding estimator δ_0 is inadmissible, outside of a closed interval. Although some progress was made later, the complete answer is still unknown.

1.2. Admissibility of $(aX + b)/(cX + d)$

In this section we consider the problem of estimating a function $g(\theta)$ which is piecewise continuous; further restrictions will be imposed later on g. To this end we first write $(aX + b)/(cX + d)$ as a formal Bayes estimator, with respect to some (generally improper) prior distribution. For more details of this kind of approach, see Zidek (1970).

If $\pi(\theta)$ is the Radon-Nikodym derivative of the prior distribution with respect to the Lebesgue measure, we can write

$$\frac{aX + b}{cX + d} = \frac{\int g(\theta)\beta(\theta)e^{\theta x}\pi(\theta)\, d\theta}{\int e^{\theta x}\beta(\theta)\pi(\theta)\, d\theta} . \tag{1.9}$$

Integrating by parts, we get

(1.10)

$$-a\int e^{\theta x}(\beta\pi)'\, d\theta + b\int e^{\theta x}\beta\pi\, d\theta = -c\int (g\beta\pi)'\, e^{\theta x}\, d\theta + d\int g\beta\pi e^{\theta x}\, d\theta$$

and, by the uniqueness of the Laplace transform:

$$(1.11) \qquad -a(\beta\pi)' + b(\beta\pi) = -c(g\beta\pi)' + d(g\beta\pi) .$$

To solve the above differential equation, we can write, after some simple calculations:

$$(1.12) \qquad (\log \beta\pi)' = \frac{dg - b}{cg - a} - (\log |cg - a|)' .$$

For simplicity, in the above formulas, we have suppressed the argument in denoting functions.

The differential equation (1.12) has the solution:

$$(1.13) \quad \pi(\theta) = \frac{1}{\beta(\theta)\ |cg(\theta) - a|} \exp \int_{\alpha}^{\theta} \frac{dg(t) - b}{cg(t) - a}\, dt \quad .$$

where α is an interior point of Θ .

We mention that the calculation above should be merely viewed as heuristic, with the goal of deriving the expression (1.13) for the prior π .

Throughout the remaining of this Chapter, we make the following assumptions:

(A1) $\quad cg(\theta) - a > 0$, for all $\theta \in \Theta$

(A2) $\quad \int_{u}^{v} \frac{dg(t) - b}{cg(t) - a}\, dt$ exists, for any $[u,v] \subset \Theta$

(A3) $\quad \int_{-\infty}^{\infty} \frac{e^{\theta x}}{(cx + d)^2}\, d\mu(x) < \infty$, for all $\theta \in \Theta$.

The main result giving sufficient conditions for admissibility is contained in the following.

Theorem 1.3. Let the density of X be $f_\theta(x) = \beta(\theta)e^{\theta x}$ and let $\underline{\theta}, \overline{\theta}$ be the endpoints of Θ. Suppose that conditions (A1) through (A3) are satisfied. Denote by

$$(1.14) \quad \sigma(\theta) = \pi(\theta)\beta(\theta)(cg(\theta) - a)^2 \int_{-\infty}^{\infty} \frac{e^{\theta x}}{(cx + d)^2} \, d\mu(x) ,$$

where $\pi(\theta)$ is given by (1.13).

If

$$(1.15) \quad \lim_{v \to \overline{\theta}} \int_u^v \sigma^{-1}(\theta) d\theta = \infty , \quad \lim_{u \to \underline{\theta}} \int_u^v \sigma^{-1}(\theta) d\theta = \infty ,$$

then $\delta_0(X) = (aX + b)/(cX + d)$ is admissible in estimating $g(\theta)$ with quadratic loss.

Proof: Suppose that $(aX + b)/(cX + d)$ is not admissible; then there exists an estimator δ, such that

$$(1.16) \int_{-\infty}^{\infty} (\delta(x) - g(\theta))^2 f_\theta(x) d\mu(x) \le \int_{-\infty}^{\infty} (\frac{ax + b}{cx + d} - g(\theta))^2 f_\theta(x) d\mu(x)$$

for all $\theta \in \Theta$ and with strict inequality for at least one $\theta \in \Theta$.

We will show that $\delta(x) = \frac{ax + b}{cx + d}$ a.e. (with respect to μ).

First, a simple calculation shows that (1.16) is equivalent to:

$$\int_{-\infty}^{\infty} (\delta(x) - \frac{ax+b}{cx+d})^2 f_\theta(x)\, d\mu(x)$$

(1.17)

$$\leq 2 \int_{-\infty}^{\infty} (\frac{ax+b}{cx+d} - \delta(x))(\frac{ax+b}{cx+d} - g(\theta))\, f_\theta(x)\, d\mu(x) .$$

Multiplying both sides by π (given by (1.13)), integrating over $[u,v] \subset \Theta$, and using Fubini's theorem, we get:

$$\int_u^v [\int_{-\infty}^{\infty} (\delta(x) - \frac{ax+b}{cx+d})^2 \beta(\theta)\, e^{\theta x}\, d\mu(x)]\pi(\theta)\, d\theta$$

(1.18)

$$\leq 2 \int_{-\infty}^{\infty} (\frac{ax+b}{cx+d} - \delta(x)) \{\int_u^v (\frac{ax+b}{cx+d} - g(\theta))\beta(\theta)\, e^{\theta x}\, \pi(\theta) d\theta\} d\mu(x) .$$

By using (1.13) and assumption (A1), the inner integral in the right hand side of (1.18) can be simplified as follows:

(1.19)

$$\int_u^v (\frac{ax+b}{cx+d} - g(\theta))\beta(\theta)\, e^{\theta x}\pi(\theta)\, d\theta$$

$$= \int_u^v (\frac{ax+b}{cx+d} - g(\theta))\, \frac{1}{(cg(\theta) - a)}\, \exp\{\theta x + \int_\alpha^\theta \frac{dg(t) - b}{cg(t) - a}\, dt\}\, d\theta$$

$$= -\frac{1}{cx+d} \int_u^v \frac{(cx+d)g(\theta) - (ax+b)}{cg(\theta) - a} \exp\{\theta x + \int_\alpha^\theta \frac{dg(t) - b}{cg(t) - a}\, dt\}\, d\theta$$

$$= -\frac{1}{cx+d} \int_u^v \frac{d}{d\theta}\, \exp\{\theta x + \int_\alpha^\theta \frac{dg(t) - b}{cg(t) - a}\, dt\}\, d\theta$$

(cont.)

$$= \frac{1}{cx+d} [\exp \{ux + \int_\alpha^u \frac{dg(t) - b}{cg(t) - a} dt\}$$

$$- \exp \{vx + \int_\alpha^v \frac{dg(t) - b}{cg(t) - a} dt\}] .$$

Denote by $T(\theta) = \int_{-\infty}^{\infty} (\delta(x) - \frac{ax+b}{cx+d})^2 \beta(\theta) e^{\theta x} d\mu(x)$;

it is enough to show that $T(\theta_0) = 0$, for some θ_0 .

By using (1.18), (1.19), and the Schwarz inequality, we get

$$\int_u^v T(\theta)\pi(\theta)\, d\theta$$

$$\leq 2 \int_{-\infty}^{\infty} (\frac{ax+b}{cx+d} - \delta(x)) \frac{1}{cx+d} \{\exp(ux + \int_\alpha^u \frac{dg(t) - b}{cg(t) - a} dt)$$

(1.20)
$$- \exp (vx + \int_\alpha^v \frac{dg(t) - b}{cg(t) - a} dt)\} d\mu(x)$$

$$\leq 2T^{\frac{1}{2}}(u)\beta^{-\frac{1}{2}}(u) \left(\int_{-\infty}^{\infty} \frac{e^{ux}}{(cx+d)^2} d\mu(x)\right)^{\frac{1}{2}} \exp\left(\int_\alpha^u \frac{dg(t) - b}{cg(t) - a} dt\right)$$

$$+ 2T^{\frac{1}{2}}(v)\beta^{-\frac{1}{2}}(v) \left(\int_{-\infty}^{\infty} \frac{e^{vx}}{(cx+d)^2} d\mu(x)\right)^{\frac{1}{2}} \exp\left(\int_\alpha^v \frac{dg(t) - b}{cg(t) - a} dt\right) =$$

(cont.)

$$= 2T^{\frac{1}{2}}(u)\beta^{\frac{1}{2}}(u)\pi(u)(cg(u) - a)\left(\int \frac{e^{ux}}{(cx+d)^2}\, d\mu(x)\right)^{\frac{1}{2}}$$

$$+ 2T^{\frac{1}{2}}(v)\beta^{\frac{1}{2}}(v)\pi(v)(cg(v) - a)\left(\int \frac{e^{vx}}{(cx+d)^2}\, d\mu(x)\right)^{\frac{1}{2}} .$$

Let us consider the following cases:

<u>Case 1</u>: $\liminf_{v \to \bar{\theta}} \pi(v)T^{\frac{1}{2}}(v)\beta^{\frac{1}{2}}(v)(cg(v) - a)\left(\int \frac{e^{vx}}{(cx+d)^2}\, d\mu(x)\right)^{\frac{1}{2}} > 0$.

By using this and (1.20), we get:

$$M(v) = \int_u^v T(\theta)\pi(\theta)d\theta \le K\pi(v)T^{\frac{1}{2}}(v)\beta^{\frac{1}{2}}(v)(cg(v) - a)$$

(1.21)

$$\times \left(\int \frac{e^{vx}}{(cx+d)^2}\, d\mu(x)\right)^{\frac{1}{2}}$$

for v in some neighborhood V of $\bar{\theta}$, and where K is a generic constant, possibly depending on u , whose exact value plays no role in the subsequent analysis. Then

(1.22) $$M(v) \le K[M'(v)\pi(v)]^{\frac{1}{2}}\beta^{\frac{1}{2}}(v)(cg(v) - a)\left(\int \frac{e^{vx}}{(cx+d)^2}\, d\mu(x)\right)^{\frac{1}{2}} ,$$

which can be written as:

$$\text{(1.23)} \qquad \frac{M'(v)}{M^2(v)} \geq \frac{1}{K\pi(v)\beta(v)(cg(v) - a)^2 \int \frac{e^{vx}}{(cx+d)^2} \, d\mu(x)} \, .$$

Choose v_1 , $v_2 \in V$, $v_1 < v_2$, and assume $M(v_1) > 0$. Then

$$\frac{1}{M(v_1)} - \frac{1}{M(v_2)} = \int_{v_1}^{v_2} \frac{M'(v)}{M^2(v)} \, dv$$

(1.24)

$$\geq \int_{v_1}^{v_2} \frac{dv}{K\pi(v)\beta(v)(cg(v) - a)^2 \left(\int \frac{e^{vx}}{(cx+d)^2} \, d\mu(x) \right)} \, .$$

Since the left-hand side is bounded by $[M(v_1)]^{-1}$, and the right-hand side equals $K^{-1}\int_{v_1}^{v_2} \sigma^{-1}(v)dv$ (where σ is given by (1.14)), we get a contradiction, by letting $v_2 \to \bar{\theta}$ and using the first part of the hypothesis (1.15).

<u>Case 2</u>: $\liminf\limits_{v \to \bar{\theta}} \pi(v) T^{\frac{1}{2}}(v) \beta^{\frac{1}{2}}(v)(cg(v) - a) \left(\int \frac{e^{vx}}{(cx+d)^2} \, d\mu(x) \right)^{\frac{1}{2}} = 0$.

Then, by using Fatou's lemma, we get

$$\int_u^{\bar\theta} T(\theta)\pi(\theta)d\theta \le 2\pi(u)T^{\frac12}(u)\beta^{\frac12}(u)(cg(u) - a)$$

(1.25)

$$\times \left(\int \frac{e^{ux}}{(cx + d)^2}\, d\mu(x)\right)^{\frac12} .$$

If we denote by $N(u) = \int_u^{\bar\theta} T(\theta)\pi(\theta)d\theta$, we can write

$$N^2(u) \le 4(-N'(u))\pi(u)\beta(u)(cg(u) - a)^2 \int \frac{e^{ux}}{(cx + d)^2}\, d\mu(x) . \tag{1.26}$$

Thus:

$$\frac{-N'(u)}{N^2(u)} \ge \frac{1}{4\pi(u)\beta(u)(cg(u) - a)^2 \left(\int \frac{e^{ux}}{(cx + d)^2}\, d\mu(x)\right)} . \tag{1.27}$$

If $N(u_0) = 0$ for some u_0 , then $T(\theta)\pi(\theta) = 0$ a.e. on $[u_0,\bar\theta]$; therefore $T(\theta_0) = 0$ for some θ_0 , and we are done.

If we assume $N(u) \neq 0$ for any u , then, by using the same argument as in Case 1, and the second half of the hypothesis (1.15), we are led to a contradiction. This ends the proof.

Remark. The assumption (A1) can be replaced throughout by $cg(\theta) - a < 0$, for all $\theta \in \Theta$.

Observe that Theorem 1.3 includes Theorem 1.1 of Karlin (1958), if we take $b = c = 0$, $d = 1$, and $g(\theta) = E_\theta X = -\beta'(\theta)/\beta(\theta)$. To obtain Theorem 1.2 of Ghosh and

Meeden (1977), take $c = 0$, $d = 1$.

As a particular case of our theorem, we give sufficient conditions for the admissibility of nonlinear estimators of the form c/X:

Corollary 1.1. *Suppose that* $g(\theta) > 0$ *for all* $\theta \in \Theta$, $\int_u^v \frac{dt}{g(t)}$ *exists for any* $[u,v] \subset \Theta$, *and* $\int_{-\infty}^{\infty} \frac{e^{\theta x}}{x^2} d\mu(x) < \infty$.

If

$$\lim_{v \to \overline{\theta}} \int_u^v [g(\theta) \int_{-\infty}^{\infty} \frac{e^{\theta x}}{x^2} d\mu(x)]^{-1} \cdot \exp(c \int_{\alpha}^{\theta} \frac{1}{g(t)} dt)\, d\theta = \infty, \tag{1.28}$$

and

$$\lim_{u \to \underline{\theta}} \int_u^v [g(\theta) \int_{-\infty}^{\infty} \frac{e^{\theta x}}{x^2} d\mu(x)]^{-1} \cdot \exp(c \int_{\alpha}^{\theta} \frac{1}{g(t)} dt)\, d\theta = \infty, \tag{1.29}$$

then $\delta_0(X) = c/X$ *is admissible in estimating* $g(\theta)$ *with quadratic loss.*

Remark. The hypotheses in the Corollary above can also be expressed in the following, equivalent form: $g(\theta)$ is positive, $1/g(\theta)$ is locally integrable in Θ, and $E_\theta(X^{-2}) < \infty$.

1.3. Examples of nonlinear admissible estimators

The examples to be presented here are related to the estimation of a function of the scale parameter in a Gamma density. Examples 1 and 2 are concerned with estimating a function of the parameter in an exponential density. A new estimator is presented in Example 2. Example 3 is concerned with the estimation of the reciprocal of the variance in an $N(0,\sigma^2)$ density. Example 4 is related to the estimation of the variance in an $N(0,\sigma^2)$ density. This example shows that the admissible estimator commonly used to estimate σ^2 is "almost inadmissible" (in a sense to be made more precise below). In Example 6 we find admissible estimators of the most general form $(aX + b)/(cX + d)$, again in the case of an exponential density.

<u>Example 1</u>. Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with exponential density $\lambda e^{-\lambda x} I_{(0,\infty)}(x)$, where $\lambda > 0$. We want to estimate $g(\lambda) = \lambda$.

Since $X = \sum_{i=1}^{n} X_i$ is a sufficient statistic for λ , we can consider only estimators based on X . The density of X is Gamma, of the form

$$(1.30) \qquad f_\lambda(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} I_{(0,\infty)}(x) \ .$$

By changing the parameter into $\theta = -\lambda$, we get:

$$(1.31) \qquad f_\theta(x) = \frac{(-\theta)^n}{\Gamma(n)} x^{n-1} e^{\theta x} I_{(0,\infty)}(x)$$

and we estimate $g(\theta) = -\theta$.

It is easy to see that conditions (1.28) and (1.29) of Corollary 1.1 are satisfied for $c = n - 2$. Thus, if $n \geq 3$, the estimator $(n-2)/X$ is admissible in estimating λ . This is a well-known result (see Ghosh and Singh (1970)).

Example 2. Consider again $X_1,\ldots,X_n$ iid with density $\lambda e^{-\lambda x} I_{(0,\infty)}(x)$, $\lambda > 0$. We want to estimate $g(\lambda) = \lambda$.

It is easy to see that if $X = \sum_{i=1}^{n} X_i$, the estimator $(n-2)/(X+k)$ is admissible in estimating λ, for any $k \geq 0$.

This result does not seem to be known. Of course, Example 1 is a particular case, for $k = 0$.

Also note that the estimators $(n-2)/(X+k)$, $k > 0$, and $(n-2)/X$ are not equivalent (i.e., the risk of $(n-2)/(X+k)$ depends on k), and, therefore, at some points $\lambda > 0$, it is possible to improve upon the risk of $(n-2)/X$.

Example 3. In this example we consider $X_1, X_2, \ldots, X_n$ normally distributed with mean 0 and variance $\sigma^2 > 0$. The function to be estimated is $1/\sigma^2$.

Since the statistic $X = \sum_{i=1}^{n} X_i^2$ is sufficient for σ^2, our admissible estimator will be a function of X.

It is well-known that $(\sum_{i=1}^{n} X_i^2)/\sigma^2$ is distributed as χ_n^2. If we denote by $\theta = -\frac{1}{2\sigma^2}$, then $\theta < 0$ and the density of X is

$$(1.32) \qquad f_\theta(x) = \frac{(-\theta)^{n/2}}{\Gamma(n/2)} e^{\theta x} x^{n/2-1} I_{(0,\infty)}(x) .$$

Also $g(\theta) = -2\theta > 0$. In looking for an admissible estimator of the form c/X , it is easily seen that conditions (1.28), (1.29) are satisfied for $c = n-4$.

Thus, if $n \geq 5$, the estimator $(n-4)/(\sum_{i=1}^{n} X_i^2)$ is admissible in estimating $1/\sigma^2$ in sampling from an $N(0,\sigma^2)$ density.

Example 4. Let us consider again $X_1, X_2, \ldots, X_n$ iid with density $N(0,\sigma^2)$ and we want to estimate $g(\sigma^2) = \sigma^2$.

If $X = \sum_{i=1}^{n} X_i^2$, it is well-known that $X/(n+2)$ is admissible in estimating σ^2 (this can be deduced easily, for example, by using Karlin's theorem 1.1).

By applying Theorem 1.3 (or, directly, Theorem 1.2), we see that $(X + k)/(n + 2)$ is admissible in estimating σ^2 , for every $k \geq 0$.

We have here a surprising property, showing that even if $X/(n+2)$ is admissible, we can strictly improve upon its risk, on "almost" the whole parameter space. In this sense we say that $X/(n+2)$ is "almost inadmissible".

To make this discussion more precise, let us denote by $Y_k = \frac{X + k}{n + 2}$, $k \geq 0$.

The risk of Y_k (with quadratic loss) can easily be computed

$$(1.33) \qquad R(Y_k,\sigma^2) = \frac{k^2 - 4\sigma^2 k + 2(n + 2)\sigma^4}{(n + 2)^2} \quad .$$

The risk of the classical estimator $Y_0 = \frac{X}{n+2}$ is $R(Y_0,\sigma^2) = \frac{2\sigma^4}{n+2}$.

Therefore $R(Y_k,\sigma^2) < R(Y_0,\sigma^2)$, if $\sigma^2 > k/4$. Roughly speaking, if k goes to 0 , then the set on which $R(Y_k,\sigma^2) < R(Y_0,\sigma^2)$ will "approach" the whole parameter space $(0,\infty)$.

For large values of σ^2 , Y_k improves substantially upon Y_0 . Also, since

$$(1.34) \quad R(Y_0,\sigma^2) - R(Y_k,\sigma^2) = -\frac{k(k-4\sigma^2)}{(n+2)^2} > -\frac{k^2}{(n+2)^2} ,$$

it follows that, for $\sigma^2 < k/4$, Y_0 does better than Y_k , but the improvement is very small (for small k) .

Thus $X/(n+2)$ is "almost inadmissible", in this sense.

<u>Example 5</u>. Suppose that $X_1, X_2, \ldots, X_n$ is a sample from the Gamma density: $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{(0,\infty)}(x)$, where $\alpha > 0$ is known, and $\beta > 0$ is unknown.

We want to estimate $g(\beta) = \beta$.

Since $X = \sum_{i=1}^{n} X_i$ is sufficient for β and the density of X is also Gamma with parameters $n\alpha$ and β , by using the same technique as in Example 1, we find that $(n\alpha - 2)/X$ is admissible for estimating β .

If $\alpha = m$ (an integer) and $n = 1$, we get the estimator $(m - 2)/X$ obtained by Ghosh and Singh (1970).

Example 6. In this example, we consider again $X_1, X_2, \ldots, X_n$ iid with exponential density $\lambda e^{-\lambda x} I_{(0,\infty)}(x)$, $\lambda > 0$, and we want to estimate $g(\lambda) = \frac{\lambda - 1}{\lambda + 1}$.

We shall find here two admissible estimators which have the most general form $(aX + b)/(cX + d)$, with $a, b, c, d \neq 0$. We denote again by $X = \sum_{i=1}^{n} X_i$, $\theta = -\lambda$. The density of X is given by (1.31) and $g(\theta) = \frac{\theta + 1}{\theta - 1}$.

We claim that if $n \geq 3$, the estimators

$$\frac{1 - X/(n-1)}{1 + X/(n-1)} , \tag{1.35}$$

$$\frac{1 - X/(n-2)}{1 + X/(n-2)} \tag{1.36}$$

are both admissible in estimating $g(\theta)$ with quadratic loss.

It is easy to see that assumptions (A1) through (A3) are satisfied. Consider the estimator (1.35), and the integral:

$$\int_{-\infty}^{\infty} \frac{e^{\theta x}}{(cx + d)^2} \, d\mu(x) = \int_{0}^{\infty} \frac{e^{\theta x}}{(x + n - 1)^2} x^{n-1} \, dx . \tag{1.37}$$

Clearly,

$$\int_{0}^{\infty} \frac{e^{\theta x}}{(x + n - 1)^2} x^{n-1} \, dx \leq \int_{0}^{\infty} x^{n-3} e^{\theta x} \, dx = \frac{\Gamma(n-2)}{(-\theta)^{n-2}} . \tag{1.38}$$

By using this inequality, it is easy to show that the hypotheses of theorem 1.3 are satisfied.

The estimator (1.36) is handled in a similar way.

1.4. Truncated parameter space

Here we give an explicit formula for an admissible estimator, in the case of truncated parameter space.

Instead of the natural parameter space Θ, we consider a subset of it, $\Theta_0 \subseteq \Theta$. The rationale is that we consider, on some *a priori* grounds, that the unknown parameter θ is restricted to belong to Θ_0.

An estimator δ_0 of $g(\theta)$ is called Θ_0-*inadmissible*, if there exists another estimator δ_1 of $g(\theta)$, such that

$$(1.39) \qquad R(g(\theta),\delta_1) \le R(g(\theta),\delta_0) \quad \text{for every} \quad \theta \in \Theta_0 ,$$

$$(1.40) \qquad R(g(\theta_0),\delta_1) < R(g(\theta_0),\delta_0) \quad \text{for some} \quad \theta_0 \in \Theta_0 .$$

An estimator δ_0 is called Θ_0-*admissible* if it is not Θ_0-inadmissible.

In general, there is no relation between Θ_0-admissibility and admissibility. In particular, an admissible estimator need not be Θ_0-admissible.

We shall consider here the particular case when $\Theta_0 = \{\theta \le \theta_0\} \subseteq \Theta$, where θ_0 is supposed to be known.

Recall that X is an observation from the one parameter exponential density.

The idea in finding a Θ_0-admissible estimator, is to use the same prior $\pi(\theta)$ given by (1.13) and to compute the generalized Bayes estimator. A simple calculation gives:

$$(1.41) \qquad \delta(x) = \frac{\int_{\Theta_0} g(\theta)\beta(\theta)\, e^{\theta x}\pi(\theta)d\theta}{\int_{\Theta_0} \beta(\theta)e^{\theta x}\, \pi(\theta)d\theta} = \frac{\int_{\underline{\theta}}^{\theta_0} g(\theta)\beta(\theta)e^{\theta x}\pi(\theta)d\theta}{\int_{\underline{\theta}}^{\theta_0} \beta(\theta)e^{\theta x}\pi(\theta)d\theta} .$$

By using the expression (1.13) of $\pi(\theta)$ we get:

$$(1.42) \qquad \delta(X) = \frac{aX + b}{cX + d} + \frac{\exp(\theta_0 + \int_{\alpha}^{\theta_0} \frac{dg(t) - b}{cg(t) - a}\, dt\,)}{(cX + d)\int_{\underline{\theta}}^{\theta_0} \frac{\exp(\theta X + \int_{\alpha}^{\theta} \frac{dg(t) - b}{cg(t) - a}\, dt)}{cg(\theta) - a}\, d\theta} .$$

In obtaining formula (1.42), we need the following fact, which is easy to prove: if $f \in L^1(\mathbb{R})$, f is absolutely continuous on any interval of $\mathbb{R}$, and $f' \in L^1(\mathbb{R})$. then $\lim_{x\to\pm\infty} f(x) = 0$.

In our case, the function

$$(1.43) \qquad f(\theta) = \exp(\theta x + \int_{\alpha}^{\theta} \frac{dg(t) - b}{cg(t) - a}\, dt)$$

satisfies these assumptions, since we assumed that the quotient in the right-hand side of (1.9) exists.

We can now give sufficient conditions for the Θ_0-admissibility of the estimator $\delta(X)$ given by (1.42):

Theorem 1.4. *Suppose, with the notations of Theorem* 1.3, *that*

$$\lim_{u \to \underline{\theta}} \int_u^v \sigma^{-1}(\theta)\,d\theta = \infty \quad . \tag{1.44}$$

Then the estimator $\delta(X)$ *given by* (1.42) *is* Θ_0-*admissible in estimating* $g(\theta)$ *with quadratic loss*.

The proof is similar to that of Theorem 1.3 and will be omitted.

Note that in proving the Θ_0-admissibility of $\delta(X)$, only the second condition in (1.15) is needed, due to the truncation of the parameter space.

Theorem 1.4 generalizes a theorem of Katz (1961) and it is also a generalization of the corresponding result of Ghosh and Meeden (1977), who found Θ_0-admissible estimators of the form $aX + b + \phi(X)$ (where $\phi(X)$ is the "correction" due to the truncation).

In the following, we give an example of admissible estimators in truncated case, when sampling from an exponential density:

Example. Consider $X_1, X_2, \ldots, X_n$ iid with exponential density $\lambda e^{-\lambda x} I_{(0,\infty)}(x)$, where $\lambda > 0$.

For the natural parameter space $\Theta = (0,\infty)$, the estimator $(n-2)/X$ is admissible in estimating $g(\lambda) = \lambda$, as in Example 1 of Section 1.3.

Suppose that we know that $\lambda \geq 1$ and we want to estimate the same function $g(\lambda) = \lambda$.

Then it is easy to see that the condition of Theorem 1.4 is satisfied, and the corresponding estimator given by (1.42) is $(1,\infty)$-admissible.

A simple calculation gives:

$$\delta(X) = \frac{n-2}{X} + \left(Xe^{X}\int_1^{\infty} t^{n-3}\, e^{-tX}\, dt\right)^{-1} . \tag{1.45}$$

An explicit formula for this estimator can be given by using

$$\int_{-\infty}^{-a} y^k e^y\, dy = (-1)^k\, k!\, e^{-a}\left[1 + \frac{a}{1!} + \dots + \frac{a^k}{k!}\right] . \tag{1.46}$$

We finally obtain

$$\delta(X) = \frac{n-2}{X} + \frac{X^{n-3}/(n-3)!}{1 + X/1! + X^2/2! + \dots + X^{n-3}/(n-3)!} . \tag{1.47}$$

In the particular case $n = 3$ (i.e., there are three observations X_1, X_2, X_3) , we get:

$$\delta(X) = \frac{1}{X} + 1 \; . \tag{1.48}$$

It is easy to compute the risk of this estimator:

$$R(\lambda,\delta) = \frac{\lambda^2 - 2\lambda + 2}{2} \; . \tag{1.49}$$

The risk of $\delta_0(X) = 1/X$ (which is admissible in the non-truncated case and $n = 3$) is

$$R(\lambda,\delta_0) = \lambda^2/2 \; . \tag{1.50}$$

We observe that $R(\lambda,\delta) < R(\lambda,\delta_0)$ for $\lambda > 1$. This shows, among other things, that δ_0 is inadmissible for the truncated problem.

CHAPTER 2

LINEAR MINIMAX ESTIMATORS

The minimax principle for estimation problems can be stated as follows. Let δ be an estimator of a function $g(\theta)$, and let $R(g(\theta),\delta)$ be the corresponding risk function. An estimator δ_0 is called minimax if

$$(2.1) \qquad \sup_{\theta\in\Theta} R(g(\theta),\delta_0) = \inf_{\delta\in\mathcal{D}} \sup_{\theta\in\Theta} R(g(\theta),\delta) ,$$

where $\mathcal{D}$ denotes the set of all estimators.

Intuitively, a minimax estimator is one which minimizes the largest possible risk. One can also say that a minimax estimator is a Bayes estimator against a prior distribution on Θ , which is least favorable for the estimation problem (see Zacks (1971), Chapter 6).

There are many relationships between minimax and admissible estimators. For example, if δ_0 is admissible and has a constant risk, then δ_0 is minimax.

In this chapter, we give sufficient conditions for the minimaxity of the classical estimator $\delta_0(X) = X$, in estimating an arbitrary (differentiable) function $g(\theta)$. Our results generalize those of Ping (1964), who gave sufficient conditions for the minimaxity of affine estimators of the form $(X + k\lambda)/(1 + \lambda)$, $\lambda > -1$ in estimating the mean $E_\theta X = m(\theta)$.

The fact that we find conditions for the minimaxity of the usual estimator $\delta(X) = X$, and we do not consider affine estimators of the most general form $aX + b$, causes no loss in generality, and is explained in Section 2.1.

Recall that X denotes a random variable, whose probability density (with respect to a σ-finite measure) belongs to the one parameter exponential family: $f_\theta(x) = \beta(\theta)\, e^{\theta x}$. As in Chapter 1, $\underline{\theta}$ and $\overline{\theta}$ denote the endpoints of the natural parameter space Θ.

In estimating an arbitrary function $g(\theta)$, the loss function which will be considered in this chapter has the form:

$$L(\delta(x)\,,\, g(\theta)) = \frac{(\delta(x) - g(\theta))^2}{\sigma^2(\theta)}\,, \tag{2.2}$$

where $\sigma^2(\theta) = \mathrm{Var}_\theta\, X = E_\theta(X - E_\theta X)^2$.

Note that, since the density of X belongs to the exponential family, we have $\sigma^2(\theta) = m'(\theta)$, where $m(\theta) = E_\theta X$.

The choice of the normalized loss (2.2) is especially desirable in those problems for which, when the loss is squared error, the minimax risk is infinite. When this happens, any estimator is minimax and the minimax principle provides no basis for choice.

In Section 2.1 we give the sufficient condition for minimaxity. The proof of Theorem 2.1 uses the Cramér-Rao inequality.

In Section 2.2 we give some examples. While Theorem 2.1 includes many of the classical minimax estimators, we concentrate on two examples where the function to be estimated is different from the mean. The presence of linear minimax estimators arises especially in estimating a function of the scale parameter in a Gamma density.

2.1. Minimaxity of X

Let $g(\theta)$ be a function of an unknown parameter θ. Later, further restrictions will be imposed on g. The risk in estimating $g(\theta)$ by $\delta(X)$ with the loss (2.2) is

$$(2.3) \quad R(\delta(X),g(\theta)) = E_\theta\{L(\delta(X),g(\theta))\} \stackrel{\text{def}}{=} R(\delta(X),\theta) .$$

The following linearity property of minimax estimators is easy to prove: $\delta(X)$ is minimax in estimating $g(\theta)$ if and only if $a\delta(X) + b$ is minimax in estimating $ag(\theta) + b$ $(a,b \in \mathbb{R}$, $a \neq 0)$.

Because of this fact, we only need to give sufficient conditions for the minimaxity of X in estimating $g(\theta)$, rather than consider more general estimators of the form $aX + b$.

Theorem 2.1. Let X have density $f_\theta(x) = \beta(\theta)\, e^{\theta x}$, and let $g(\theta)$ be a differentiable function of θ. Denote the endpoints of Θ by $\underline{\theta}$ and $\overline{\theta}$, respectively. Suppose that the following conditions are satisfied:

(i) $\sup_{\theta \in \Theta} \frac{(g(\theta) - m(\theta))^2}{\sigma^2(\theta)} = \lim_{\theta \to \bar{\theta}} \frac{(g(\theta) - m(\theta))^2}{\sigma^2(\theta)} < \infty$,

(ii) $\liminf_{\theta \to \bar{\theta}} \beta^{-1}(\theta)\sigma(\theta) \exp(-\int_a^\theta g(t)dt) > 0$,

(iii) $\lim_{\theta \to \bar{\theta}} \int_a^\theta \beta(t) \exp(\int_a^t g(s)\, ds)\, dt = \infty$,

where $a \in \operatorname{Int} \Theta$.

Then X is minimax in estimating $g(\theta)$ with loss (2.2).

Proof: We shall use the Cramér-Rao inequality. If $\delta(X)$ is an estimator of $g(\theta)$ with bias function $b_\delta(\theta) = E_\theta(\delta(X)) - g(\theta)$, then

$$\text{(2.4)} \qquad \operatorname{Var}_\theta \delta(X) \geq \frac{(b'(\theta) + g'(\theta))^2}{I_X(\theta)}$$

where $I_X(\theta) = E_\theta(\frac{\partial}{\partial\theta} \log f_\theta(X))$.

From (2.4), since $I_X(\theta) = \sigma^2(\theta)$ for the exponential family, we get

$$\text{(2.5)} \quad R(\delta(X), g(\theta)) > \frac{1}{\sigma^2(\theta)} [b^2(\theta) + \frac{(b'(\theta) + g'(\theta))^2}{\sigma^2(\theta)}] .$$

For the rest of the proof, for simplicity, we suppress the variables in denoting functions.

Suppose that X is not minimax, i.e.,

$$\sup_\theta R(X,\theta) > \inf_\delta \sup_\theta R(\delta(X),\theta) \quad (2.6.)$$

First, note that the risk of X is given by

$$R(X,\theta) = 1 + \frac{(g(\theta) - m(\theta))^2}{\sigma^2(\theta)} \quad (2.7)$$

Thus, from (2.5), (2.6), and (2.7), there exists an $\varepsilon > 0$ and an estimator δ such that

$$\frac{1}{\sigma^2}[b^2 + \frac{(b' + g')^2}{\sigma^2}] < 1 + \sup_\theta \frac{(g(\theta) - m(\theta))^2}{\sigma^2(\theta)} - \varepsilon . \quad (2.8)$$

Let $K = \sup_\theta \frac{(g(\theta) - m(\theta))^2}{\sigma^2(\theta)}$, and let

$$u(\theta) = b(\theta) + (g(\theta) - m(\theta)) . \quad (2.9)$$

Since $b' + g' = m' + u'$, (2.9) becomes

$$u^2 + (g - m)^2 - 2(g - m)u + \frac{(m' + u')^2}{\sigma^2} < (1 + K - \varepsilon)\sigma^2 . \quad (2.10)$$

We now relax the inequality (2.10)(i.e., neglect the term u'^2) to obtain (note that $m' = \sigma^2$):

$$u^2 - 2(g - m)u + 2u' < -\varepsilon\sigma^2 + \sigma^2\{K - \frac{(g - m)^2}{\sigma^2}\} . \quad (2.11)$$

By using the assumption (i), there exists a point $\theta_1 \in \Theta$, such that for all $\theta_1 < \theta < \overline{\theta}$, we have $K - \frac{(g - m)^2}{\sigma^2} < \frac{\varepsilon}{2}$. Thus:

$$(2.12) \qquad u^2 - 2(g - m)u + 2u' < -\frac{\varepsilon}{2}\sigma^2$$

for all $\theta > \theta_1$.

Define a new function $v = \beta^{-1} \exp(-\int_a^\theta g(t)\,dt)\, u$. Then, after some calculations, (2.12) simplifies to

$$(2.13) \qquad \beta^2 \exp(2\int_a^\theta g(t)dt)v^2 + 2\beta\exp(\int_a^\theta g(t)dt)v' < -\frac{\varepsilon}{2}\sigma^2 ,$$

which can be written as

$$(2.14) \qquad \beta \exp(\int_a^\theta g(t)dt)v^2 + 2v' < -\frac{\varepsilon}{2}\sigma^2\beta^{-1} \exp(-\int_a^\theta g(t)dt)$$

for all $\theta > \theta_1$.

Now, by (ii), there exists a constant $c > 0$ and a point θ_1', such that for all $\theta > \theta_1'$, we have $\beta^{-1}\sigma \exp(-\int_a^\theta g(t)dt) > c$.

Then, for $\theta > \theta_1'' = \max(\theta_1, \theta_1')$, we have by (2.14):

$$(2.15) \qquad \beta\exp(\int_a^\theta g(t)dt)v^2 + 2v' < -\frac{\varepsilon}{2}\, c^2\beta\exp(\int_a^\theta g(t)dt) .$$

Note that $v' < 0$ for $\theta > \theta_1''$, i.e., v decreases for $\theta > \theta_1''$, and so $\lim_{\theta \to \overline{\theta}} v(\theta)$ exists. We get

$$(2.16) \qquad 2v'[v^2 + \frac{\varepsilon c^2}{2}]^{-1} < -\beta \exp\left(\int_a^\theta g(t)dt\right)$$

for all $\theta > \theta_1''$.

Integrating both sides of (2.16), we obtain

$$(2.17) \quad \frac{2}{c}\sqrt{\frac{2}{\varepsilon}} \tan^{-1}\left(\sqrt{\frac{2}{\varepsilon}} \cdot \frac{v}{c}\right)\Bigg|_{\theta_1''}^{\theta} \le -\int_{\theta_1''}^{\theta} \beta(t) \exp\left(\int_a^t g(s)ds\right) dt .$$

Letting $\theta \to \overline{\theta}$, the right-hand side of (2.17) approaches $-\infty$ by assumption (iii), while the left-hand side is finite. This contradiction ends the proof of the theorem.

Remarks. (a) If we let $h = \frac{g - m}{\sigma}$, then, since $\exp\left(\int_a^\theta m(t)dt\right) = d\beta^{-1}(\theta)$, where d is a positive constant, Theorem 2.1 can be restated as follows:

Assume

(i)' $\sup_\theta h^2(\theta) = \lim_{\theta \to \overline{\theta}} h^2(\theta) < \infty$,

(ii)' $\liminf_{\theta \to \overline{\theta}} \sigma \exp\left(-\int_a^\theta \sigma h(t)dt\right) > 0$,

(iii)' $\int_a^{\overline{\theta}} \exp\left(\int_a^\theta \sigma h(t)dt\right)d\theta = \infty$.

Then X is minimax in estimating $\sigma h + m$ with loss (2.2).

If $h = 0$, and $\Theta = \mathbb{R}$, then Theorem 2.1 takes a very simple form:

If

$$\text{(2.18)} \qquad \liminf_{\theta \to +\infty} \sigma^2(\theta) > 0 \quad \text{or} \quad \liminf_{\theta \to -\infty} \sigma^2(\theta) > 0 ,$$

then X is minimax in estimating $m(\theta)$.

Indeed, in this case (i)' and (iii)' are obviously satisfied, while (ii)' (or its dual form) becomes (2.18) above.

We point out, however, that this condition (ii)', while being sufficient, is not necessary for the minimaxity of X. Indeed, consider a random variable X whose distribution is binomial with parameters $(n , \frac{e^\theta}{1 + e^\theta})$, where $\theta \in \Theta = (-\infty,\infty)$. Then

$$\sigma^2(\theta) = \frac{ne^\theta}{(1 + e^\theta)^2} , \text{ and } \lim_{\theta \to \pm\infty} \sigma^2(\theta) = 0 .$$

However, it follows from Ghirschick and Savage (1951) that X is admissible and, having a constant risk, it is minimax.

(b) Our condition (i) is more general than a similar condition of Ping (1964), and also (ii) allows for the limit inferior to be infinite (this is, actually, the case in many examples).

(c) It is interesting to compare conditions (i) - (iii) with the sufficient conditions for admissibility (Theorem 1.2). While the latter involve the behavior of a certain integral in the neighborhood of both endpoints $\underline{\theta}$ and $\overline{\theta}$, the former involve the behavior of the same integral and of another expression in the neighborhood of one endpoint ($\bar{\theta}$, say), and also a global condition, (i). This global condition is a kind of maximum principle for the square of the normalized function $h(\theta) = \frac{g(\theta) - m(\theta)}{\sigma(\theta)}$.

2.2. Examples

It is easy to see that Theorem 2.1 can be used to prove the minimaxity of many classical estimators. For example, if $X_1, X_2, \ldots, X_n$ are normally distributed with mean μ and variance 1, then, by using Theorem 2.1 (and simple changes of scale), it is easy to check that

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{2.19}$$

is minimax in estimating the mean μ with quadratic loss.

Some more examples of this kind are given in Ping (1964) although his condition analogous to (i) needs some special care.

We now describe two examples which are related to the estimation of a function of the scale parameter in a Gamma density.

Example 1. Suppose that $X_1, X_2, \ldots, X_n$ are iid with exponential density $\lambda e^{-\lambda x} I_{(0,\infty)}(x)$ where $\lambda > 0$. We want to estimate the function

$$g(\lambda) = \frac{n + e^{-\lambda}}{\lambda} \tag{2.20}$$

Since $X = \sum_{i=1}^{n} X_i$ is a sufficient statistic for λ, we can consider estimators based on X. The density of X is Gamma with parameters n and λ.

By changing the parameter to $\theta = -\lambda$, we get

$$(2.21) \qquad f_\theta(x) = \frac{(-\theta)^n}{\Gamma(n)} x^{n-1} e^{\theta x} I_{(0,\infty)}(x) , \quad \theta < 0 .$$

We claim that X is minimax in estimating $g(\theta) = -\frac{n + e^\theta}{\theta}$, with the loss given by (2.2).

It is easy to see that the assumptions (i)'-(iii)' are satisfied with $h = e^\theta/\sqrt{n}$. Expand e^t/t as a power series: $e^t/t = \frac{1}{t} + 1 + \frac{t}{2!} + \frac{t^2}{3!} + \ldots$. Then

$$(2.22) \qquad \int_a^\theta \frac{e^t}{t}\, dt = \log(-\theta) + \sum_{k=1}^{\infty} \frac{\theta^k}{k!k} + c ,$$

where c is a constant. Thus,

$$(2.23) \qquad \liminf_{\theta\to 0} \sigma \exp\left(-\int_a^\theta \sigma h(t)dt\right) = \sqrt{n}\, e^c \lim_{\theta\to 0} \exp\left(\sum_{k=1}^{\infty} \frac{\theta^k}{k!k}\right) > 0 ,$$

and so (ii)' is satisfied.

Also, since $\lim_{\theta\to 0} \exp\left(\sum_{k=1}^{\infty} \frac{\theta^k}{k!k}\right) = 1$, there exists $\theta_1 < 0$ such that for all $\theta_1 < \theta < 0$, we have

$$(2.24) \qquad \int_a^0 \exp\left(\int_a^\theta \sigma h(t)dt\right)d\theta > \int_a^0 \left(-\frac{1}{2\theta}\right)d\theta = \infty$$

showing that (iii)' is satisfied. Finally, since $h = \frac{e^\theta}{\sqrt{n}}$ and $\bar{\theta} = 0$, (i)' is also true and the claim is proved.

Example 2. In this example we consider $X_1, X_2, \ldots, X_n$ normally distributed with mean 0 and variance $\sigma^2 > 0$. The function to be estimated is $(n + 2)\sigma^2 - \frac{2\sigma^2}{2\sigma^2 + 1}$.

Since $X = \sum_{i=1}^{n} X_i^2$ is sufficient for σ^2, we can restrict our attention to estimators based on X. It is well-known that $(\sum_{i=1}^{n} X_i^2)/\sigma^2$ is χ_n^2. If we denote by $\theta = -\frac{1}{2\sigma^2}$, then $\theta < 0$ and the density of X is

$$(2.25) \qquad f_\theta(x) = \frac{(-\theta)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{\theta x} I_{(0,\infty)}(x) .$$

Also $m(\theta) = -\frac{n}{2\theta}$, $\sigma^2(\theta) = \frac{n}{2\theta^2}$, and

$$(2.26) \qquad g(\theta) = [(-\tfrac{n}{2} - 1)/\theta] - [1/(1 - \theta)] .$$

With $h = \sqrt{\frac{2}{n}}\Big/(1 - \theta)$, it is easily seen that conditions (i)' - (iii)' are satisfied.

Thus X is minimax in estimating $(n + 2)\sigma^2 - \frac{2\sigma^2}{2\sigma^2 + 1}$.

CHAPTER 3

IMPROVING UPON THE BEST INVARIANT ESTIMATOR IN MULTIVARIATE LOCATION PROBLEMS

In this chapter we consider the problem of finding estimators which are better than the best invariant estimator $\delta_o(X) = X$, where X is a p-dimensional random vector whose probability density belongs to the location family: $f_\theta(x) = f(x - \theta)$, $\theta \in \mathbb{R}^p$.

We give sufficient conditions for the inadmissibility in dimensions $p \geq 3$ of the best invariant estimator $\delta_o(X) = X$, in estimating the location parameter θ with convex loss function $L(\theta, \delta(x)) = L(\delta(x) - \theta)$. In the particular case when the loss function is

$$(3.1) \qquad L(\theta, \delta(x)) = \sum_{i=1}^{p} c_i(\delta_i(x) - \theta_i)^2$$

where $c_1, c_2, \ldots, c_p$ are given positive constants, and if we make suitable assumptions about the moments of the density $f(x - \theta)$, we derive various classes of estimators which improve upon the best invariant estimator $\delta_o(X) = X$.

Our problem originates from Brown (1975) who proved that in dimensions $p \geq 3$, the estimator $\delta_o(X) = X$ is inadmissible in estimating the mean θ of a multivariate normal distribution (with covariance matrix the identity), under the loss function (3.1).

In Section 3.1 we generalize a result of Brown (1975) concerning sufficient conditions for inadmissibility. As appropriate background we give a result of Brown (1966) concerning the inadmissibility in dimensions $p \geq 3$ of the best invariant estimator $\delta_o(X) = X$ in estimating the location parameter θ with loss function (3.1). In Section 3.2 we prove that under suitable assumptions, the convolution of an estimator δ_1 (which improves upon the best invariant estimator $\delta_o(X) = X$ outside of a compact set) with a truncated probability density in $\mathbb{R}^p$, gives an estimator δ which is uniformly better than δ_o. The estimator δ_1 is of the type $\delta_1(X) = (1 - \frac{a}{||X||^2})X$, where a is a suitable constant (such estimators are called James-Stein estimators). We also give several examples of estimators obtained in this way, which improve upon δ_o. In Section 3.3 we derive some other estimators which improve upon δ_o in higher dimensions. The critical dimension for which we have an improvement upon δ_o depends on δ_1, the latter being an estimator which improves upon δ_o outside of a compact set and which is not of the James-Stein type.

Finally, we state some problems which were not solved in the present context and which could be of further research interest.

3.1. The inadmissibility result

Suppose the density of X to be of the location type $f(x - \theta)$ and consider the general loss function $L(\theta, \delta(x)) = L(\delta(x) - \theta)$.

By $||\cdot||$ we denote the Euclidean norm in $\mathbb{R}^p$ and $[q]$ denotes the largest integer not exceeding q .

The key lemma which is proved below gives sufficient conditions for the inadmissibility of δ_o . It is a generalization of proposition 1 in Brown (1975).

Lemma 3.1. Suppose that the following hypotheses are satisfied:

(i) L is a convex function;

(ii) there exists an estimator δ_1 , whose risk $R(\theta,\delta_1)$ is bounded on compact subsets of $\mathbb{R}^p$;

(iii) $\liminf_{\|\theta\|\to\infty} \|\theta\|^q [R(\theta,\delta_o) - R(\theta,\delta_1)] > 0$, for some $q > 0$.

Then, for $p \geq [q] + 1$, the estimator $\delta_o(X) = X$ is inadmissible in estimating θ with loss function $L(\delta(x) - \theta)$.

Proof: Denote by $\Delta(\theta) = R(\theta,\delta_o) - R(\theta,\delta_1)$ and let $0 < a < \liminf_{\|\theta\|\to\infty} \|\theta\|^q \Delta(\theta)$. Then, for some $r > 0$, we have $\Delta(\theta) \geq a/\|\theta\|^q$, for $\|\theta\| > r$.

Obviously $R(\theta,\delta_o) = R_o$ (a constant) and, by (ii), $R(\theta,\delta_1) \leq B$ for $\|\theta\| \leq r$ (where B denotes a suitable constant). Thus

$$\Delta(\theta) \geq \phi(\|\theta\|) \tag{3.2}$$

where the function ϕ is defined by:

$$\phi(t) = \begin{cases} R_o - B \ , & \text{if } 0 \leq t \leq r \\ a/t^q \ , & \text{if } t > r \ . \end{cases} \tag{3.3}$$

Now we use the "randomization of the origin" argument of Brown (1975): denote by

$$\delta_1^\tau(x) = \tau + \delta_1(x - \tau) \quad , \tau \in \mathbb{R}^p . \tag{3.4}$$

Observe that $R(\theta,\delta_1^\tau) = R(\theta - \tau,\delta_1)$, since θ is a location parameter.

The idea is now to consider τ as a random variable (whose distribution will be specified later) and, by taking

$$\delta_2(X) = E_\tau[\delta_1^\tau(X)] \tag{3.5}$$

to try to improve upon the risk of δ_o .

A calculation using the fact that L is a convex function and applying the Jensen inequality, gives:

$$R(\theta,\delta_2) \le E_\tau[R(\theta,\delta_1^\tau)] = E_\tau[R(\theta - \tau,\delta_1)] \tag{3.6}$$

Note that $R(\theta,\delta_o) = E_\tau[R(\theta - \tau,\delta_o)]$, since we have $R(\theta - \tau,\delta_o) = R(\theta,\delta_o^\tau) = R(\theta,\delta_o)$. Thus:

$$\begin{aligned} R(\theta,\delta_o) - R(\theta,\delta_2) &\ge E_\tau[R(\theta - \tau,\delta_o) - R(\theta - \tau,\delta_1)] \\ &= E_\tau[\Delta(\theta - \tau)] \ge E_\tau[\phi(\|\theta - \tau\|)] \end{aligned} \tag{3.7}$$

where E_τ denotes the integral with respect to the probability measure associated to the random variable τ .

Now, choose τ uniformly distributed in the ball with center 0 and radius K in $\mathbb{R}^p$:

$$B_K(0) = \{z \in \mathbb{R}^p \,/\, \|z\| \le K\} . \tag{3.8}$$

We show that a suitable choice of $K > 0$ will imply that

$$E_\tau[\phi(\|\theta - \tau\|)] > 0 \text{ , for all } \theta \in \mathbb{R}^p . \tag{3.9}$$

This will prove the inadmissibility of δ_o .

We have:

$$E[\phi(\|\theta - \tau\|)] = \frac{1}{\alpha K^p}\left[\int_{\substack{\|\tau\| \le K \\ \|\theta-\tau\| \le r}} (R_o - B)\, d\tau + \int_{\substack{\|\tau\| \le K \\ \|\theta-\tau\| > r}} \frac{a}{\|\theta - \tau\|^q}\, d\tau \right] \tag{3.10}$$

where $\alpha = K^{-p}$ Volume $B_K(0)$.

Suppose that $R_o - B < 0$ since, otherwise, we are done. Observe that if $\|\theta\| > K + r$, then:

$$E[\phi(\|\theta - \tau\|)] = \frac{1}{\alpha K^p} \int_{\|\tau\| \le K} \frac{a}{\|\theta - \tau\|^q}\, d\tau > 0 . \tag{3.11}$$

So, it remains to find $K > 0$, such that, for $\|\theta\| \le K + r$, $E[\phi(\|\theta - \tau\|)] > 0$.

But, if λ_p denotes the p-dimensional Lebesgue measure, we have:

$$(3.12) \qquad E[\phi(\|\theta - \tau\|)] \geq \frac{1}{\alpha K^p}\left[(R_o - B)\lambda_p(A) + \frac{a}{(2K + r)^q}\lambda_p(A_1)\right]$$

where $A = \{\tau \mid \|\tau\| \leq K\} \cap \{\tau \mid \|\tau - \theta\| \leq r\}$ and $A_1 = \{\tau \mid \|\tau\| \leq K\} \setminus A$. Since

$$(3.13) \qquad \lambda_p(A) + \lambda_p(A_1) = \lambda_p(B_K(0)) = \alpha K^p$$

we can write:

$$(3.14) \qquad E[\phi(\|\theta - \tau\|)] \geq \frac{1}{\alpha K^p}\left[(R_o - B)\lambda_p(A) + \frac{a}{(2K + r)^q}(\alpha K^p - \lambda_p(A))\right] .$$

Clearly $\lambda_p(A) \leq \alpha r^p$ and we get:

$$(3.15) \qquad \begin{aligned} &E[\phi(\|\theta - \tau\|)] \\ &\geq \frac{1}{\alpha K^p}\left\{\left[R_o - B - \frac{a}{(2K + r)^q}\right]\alpha r^p + \frac{a\alpha K^p}{(2K + r)^q}\right\} \end{aligned} .$$

Since for $p \geq [q] + 1$ we have:

$$(3.16) \qquad \lim_{K\to\infty}\left\{\left[R_o - B - \frac{a}{(2K + r)^q}\right]\alpha r^p + \frac{a\alpha K^p}{(2K + r)^q}\right\} = \infty$$

the proof of the inadmissibility of δ_o is completed.

Note. Observe that condition (ii) in the statement of Lemma 3.1 is more general than the corresponding one in Brown (1975) where $R(\theta, \delta_1)$ is required to be bounded on $\mathbb{R}^p$.

In particular, our condition (ii) is satisfied whenever

$R(\theta, \delta_1)$ is a continuous function of θ .

From now on, unless otherwise stated, we shall consider the more particular loss function (3.1). Recall that X has a density of the location type $f(x - \theta)$.

Denote by $Z = X - \theta$; obviously Z has the density $f(x)$ which is independent of θ .

We shall make the following assumptions:

(a) $E(Z) = (E(Z_1), E(Z_2), \ldots, E(Z_p)) = 0$

(b) $E(Z_i Z_j) = 0$, for all $i \neq j$.

Clearly, (a) is a mild assumption, which is necessary to show that the best invariant estimator of θ is $\delta_o(X) = X$.

Assumption (b) combined with (a) states that Z_i and Z_j are uncorrelated random variables, for every $i \neq j$.

We shall now express the risk difference $R(\theta, \delta_o) - R(\theta, \delta)$, where δ is any estimator of θ , in a special form. This calculation, similar to the one in Brown (1975), is of fundamental importance for all that follows.

Let δ be any estimator; we have:

$$\begin{aligned} \Delta(\theta) = R(\theta, \delta_o) - R(\theta, \delta) &= \sum_{i=1}^{p} c_i [E(X_i - \theta_i)^2 - E(\delta_i(X) - \theta_i)^2] \\ &= \sum_{i=1}^{p} c_i \Delta_i(\theta) \end{aligned} \tag{3.17}$$

and

$$\Delta_i(\theta) = E(X_i - \delta_i(X))(X_i + \delta_i(X)) - 2\theta_i E(X_i - \delta_i(X)) .$$

Let $h(X) = X - \delta(X)$ and $Z = X - \theta$; then:

$$(3.18)\qquad \begin{aligned} \Delta_i(\theta) &= Eh_i(Z+\theta)[2Z_i - h_i(Z+\theta)] \\ &= 2E[Z_i h_i(Z+\theta)] - E[h_i^2(Z+\theta)] \ . \end{aligned}$$

By using Taylor's formula, we get:

$$(3.19)\qquad h_i(Z+\theta) = h_i(\theta) + \sum_j Z_j h_{ij}(\theta) + e_i(\theta, Z)$$

where $h_{ij}(x) = \frac{\partial h_i}{\partial x_j}(x)$ and e_i is an error term. Therefore (3.18) becomes:

$$(3.20)\qquad \Delta_i(\theta) = 2\alpha_i h_i(\theta) + 2\sum_j a_{ij} h_{ij}(\theta) - E[h_i^2(Z+\theta)] + e_i'(\theta)$$

where $\alpha_i = E(Z_i)$, $a_{ij} = E(Z_i Z_j)$, $e_i'(\theta) = 2E(Z_i e_i)$. Note that α_i and a_{ij} do not depend on θ , and also that, for the sake of generality, in this calculation we do not assume the conditions (a) and (b) above.

By using again (3.19) and by rewriting the error term, we get:

$$(3.21)\qquad \Delta_i(\theta) = 2\alpha_i h_i(\theta) + 2\sum_j a_{ij} h_{ij}(\theta) - h_i^2(\theta) + e_i''(\theta) \ .$$

Finally, by using (3.17), we obtain:

$$(3.22)\qquad \Delta(\theta) = 2\sum_i c_i \alpha_i h_i(\theta) + 2\sum_{i,j} c_i a_{ij} h_{ij}(\theta) - \sum_i c_i h_i^2(\theta) + e''(\theta) \ .$$

This formula can be written as:

$$\Delta(\theta) = D(\theta) + e''(\theta) \tag{3.23}$$

where

$$D(\theta) = 2\sum_i c_i \alpha_i h_i(\theta) + 2\sum_{i,j} c_i a_{ij} h_{ij}(\theta) - \sum_i c_i h_i^2(\theta) \; . \tag{3.24}$$

Now, we can state the main result of this section. In the following theorem and whenever we shall use this theorem, we shall assume that the density $f(x - \theta)$ satisfies certain moment properties as described in detail in Brown (1966).

Theorem 3.1. Suppose that X has density $f(x - \theta)$, $\theta \in \mathbb{R}^p$. If the following conditions are satisfied:

(a) $E_o(X) = (E_o(X_1), E_o(X_2), \ldots, E_o(X_p)) = 0$

(b) $E_o(X_i X_j) = 0$ for all $i \neq j$

(c) $p \geq 3$

then the estimator $\delta_o(X) = X$ is inadmissible in estimating θ with loss function given by (3.1).

Proof: Note that in (a) and (b), E_o denotes the expected value when $\theta = 0$ (these hypotheses are the same as (a) and (b) discussed before).

We shall use the calculation above. A discussion of the error terms appears in Brown (1975); a more complete discussion is given in Brown (1966). Since these apply to our case, we do not repeat these calculations concerning the errors here.

From formula (3.23) and from Lemma 3.1, it is enough to show that

$$\liminf_{\|\theta\|\to\infty} \|\theta\|^2 D(\theta) > 0 \tag{3.25}$$

for a suitable estimator $\delta(X)$.

Clearly, from assumptions (a) and (b), we have:

$$D(\theta) = 2\sum_i c_i a_{ii} h_{ii}(\theta) - \sum_i c_i h_i^2(\theta) \tag{3.26}$$

where $a_{ii} = E_o(X_i^2) = E(Z_i^2) = \text{Var}\,(Z_i)$ does not depend on θ .

Consider the estimator $\delta_i(X) = X_i - \dfrac{\varepsilon X_i}{c_i a_{ii}\|X\|^2}$, for $\|X\| \geq 1$. Then $h_{ii}(\theta) = \dfrac{\varepsilon}{c_i a_{ii}} \cdot \dfrac{\|\theta\|^2 - 2\theta_i^2}{\|\theta\|^4}$. Therefore:

$$\begin{aligned} D(\theta) &= 2\sum c_i a_{ii} \frac{\varepsilon}{c_i a_{ii}} \frac{\|\theta\|^2 - 2\theta_i^2}{\|\theta\|^4} - \sum c_i \frac{\varepsilon^2\theta_i^2}{c_i^2 a_{ii}^2 \|\theta\|^4} \\ &= \frac{\varepsilon}{\|\theta\|^2} [2p - 4 - \sum_i \frac{\varepsilon\theta_i^2}{c_i a_{ii}^2 \|\theta\|^2}] , \quad \|\theta\| > 1 . \end{aligned} \tag{3.27}$$

Now, since $\sum_i \dfrac{\theta_i^2}{\|\theta\|^2} = 1$, if we choose

$$0 < \varepsilon < 2(p-2) \min_{1\leq i\leq p} (c_i a_{ii}^2) \tag{3.28}$$

and, since $p \geq 3$ by (c), it follows that

$$(3.29) \qquad \|\theta\|^2 D(\theta) > \varepsilon[2p - 4 - \frac{\varepsilon}{\min_i c_i a_{ii}^2}] > 0 .$$

Finally, $\liminf_{\|\theta\| \to \infty} \|\theta\|^2 \Delta(\theta) \geq \varepsilon(2p - 4 - \frac{\varepsilon}{\min_i c_i a_{ii}^2}) > 0$

and, thus, δ_o is inadmissible.

Note. We mention that Theorem 3.1 generalizes the original result of Stein (1955) and it is included in the more general inadmissibility result of Brown (1966). The point here is the possibility that the components of X might be dependent (but uncorrelated) and that we used the uniform distribution in Lemma 3.1. As we shall see in the next section, the latter argument leads to possibilities of generalizations and to a wide class of estimators which improve upon the best invariant procedure δ_o.

We now give some examples where Theorem 3.1 can be applied:

Example 1. Consider X normally distributed with mean θ and covariance matrix $\sigma^2 I$, where σ^2 is known. In this case, obviously, conditions (a) and (b) are satisfied. Thus, if $p \geq 3$, the estimator $\delta_o(X) = X$ is inadmissible. This is the classical problem of estimating the mean vector of a multivariate normal distribution.

Example 2. Consider X to be uniformly distributed in the ball $\{x \in \mathbb{R}^p / \|x - \theta\| \leq R\}$. In this case we have:

$$(3.30) \qquad f(x) = \frac{1}{\alpha R^p} I_{\{\|x\| \leq R\}}$$

and, by using polar coordinates, it can be easily shown that (a) and (b) are satisfied. This is the case of a spherically uniform distribution. Minimax estimators, better than δ_o, were obtained in this case by Brandwein and Strawderman (1978).

Example 3. Consider the observation X having a density of the form $f(\|x - \theta\|)$, $\theta \in \mathbb{R}^p$. It can be shown, again by using polar coordinates, that if

$$\int_o^\infty t^{p+1} f(t)\, dt < \infty \tag{3.31}$$

then conditions (a) and (b) are satisfied. In particular, any truncated density (see the next section) will satisfy (3.31) and, therefore, when sampling from such a density, the estimator $\delta_o(X) = X$ is inadmissible, for $p \geq 3$.

3.2. Improving upon the best invariant estimator

The problem of improving upon the best invariant estimator $\delta_o(X) = X$ received considerable attention in the literature, but estimators better than δ_o are only known in special cases, such as in sampling from a multivariate normal density, or from a spherically symmetric density.

In the general case of a location parameter, let us observe that Lemma 3.1 gives a family of estimators which improve upon δ_o in terms of risk:

$$\delta_2(X) = \frac{1}{\alpha K^p} \int_{\|\tau\| \leq K} [\tau + \delta_1(X - \tau)]\, d\tau \;, \quad K \geq K_o \;. \tag{3.32}$$

If we note that $\int_{\|\tau\| \le K} \tau \, d\tau = 0$ (by using polar coordinates in $\mathbb{R}^p$), we obtain:

$$(3.33) \qquad \delta_2(X) = \frac{1}{\alpha K^p} \int_{\|\tau\| \le K} \delta_1(X - \tau) d\tau \ , \ K \ge K_o$$

where δ_1 is an estimator satisfying the conditions of Lemma 3.1.

Formula (3.33) shows that the estimator δ_2 is obtained as the convolution of an estimator δ_1 which improves outside of a compact set, with a suitable uniform density in $\mathbb{R}^p$.

The possibility of expressing estimators which improve upon δ_o as a convolution, will be taken as the basis of further developments. An important step in finding wider classes of estimators which are better than δ_o , is to generalize Lemma 3.1. We do this below and, basically, this generalization shows that a wide class of densities can be taken instead of a uniform density, to generate estimators which improve upon δ_o .

We shall consider <u>truncated distributions</u>, whose densities (with respect to the Lebesgue measure) are of the form:

$$(3.34) \qquad \xi(\|x\|^2) = \begin{cases} c \ \eta(\|x\|^2) & , \text{ if } \|x\| \le K \\ 0 & , \text{ if } \|x\| > K \end{cases}$$

where $K > 0$ and

$$(3.35) \qquad c^{-1} = \int_{\|x\| \le K} \eta(\|x\|^2) dx \ .$$

The following theorem generalizes Lemma 3.1. Note that we consider again the general loss function $L(\theta, \delta(\mathbf{x})) = L(\delta(\mathbf{x}) - \theta)$.

Theorem 3.2. Suppose the following hypotheses satisfied:

(i) L is a convex function;

(ii) there exists an estimator δ_1 , whose risk $R(\theta, \delta_1)$ is bounded on compact sets;

(iii) $\liminf_{\|\theta\| \to \infty} \|\theta\|^2 [R(\theta, \delta_o) - R(\theta, \delta_1)] > 0$;

(iv) $\sup_{y \in \mathbf{R}^p} \int_{\|x-y\| \le r} \eta(\|x\|^2)\, dx < \infty$, for any $r > 0$;

(v) $\lim_{K\to\infty} (1/K^2) \int_{\|x\| \le K} \eta(\|x\|^2) dx = \infty$.

Then, the estimator defined by

$$(3.36) \qquad \delta_2(X) = c \int_{\|\tau\| \le K} \delta_1(X - \tau) \eta(\|\tau\|^2) d\tau$$

is better than the best invariant estimator δ_o , for $K \ge K_o$ (where K_o is a sufficiently large constant).

Proof: The method of proof is similar to that of Lemma 3.1 and, by using the same notations, we get:

$$(3.37) \qquad R(\theta, \delta_o) - R(\theta, \delta_2) \ge E_\tau[\phi(\|\theta - \tau\|)] .$$

We choose the random variable τ distributed with density ξ as given by (3.34), so that:

$$R(\theta,\delta_o) - R(\theta,\delta_2)$$

(3.38)
$$\geq c\Big[\int_{\substack{\|\theta-\tau\|\leq r\\ \|\tau\|\leq K}} (R_o - B)\eta(\|\tau\|^2)d\tau + \int_{\substack{\|\theta-\tau\|>r\\ \|\tau\|\leq K}} \frac{a}{\|\theta-\tau\|^2}\eta(\|\tau\|^2)d\tau\Big].$$

Consider $\|\theta\| < K + r$ since, otherwise, we are done. We obtain:

$$R(\theta,\delta_o) - R(\theta,\delta_2)$$

(3.39)
$$\geq c\Big\{\Big[R_o - B - \frac{a}{(2K+r)^2}\Big]\int_{\|\tau-\theta\|\leq r} \eta(\|\tau\|^2)d\tau + \frac{a}{c(2K+r)^2}\Big\}.$$

Finally, by using hypotheses (iv) and (v), we show that the right-hand side of (3.39) goes to ∞ as $K \to \infty$, which concludes the proof.

Observe that δ_2 is obtained from δ_1 by a randomization of the origin. Since:

(3.40)
$$\int_{\|\tau\|\leq K} \tau\eta(\|\tau\|^2)d\tau = 0$$

we get formula (3.36), providing an estimator δ_2 with smaller risk than δ_o.

Again, the estimator δ_2 is the convolution of the estimator δ_1 (which improves upon δ_o outside of a compact set) with a truncated density of the type (3.34).

We give now some examples which, according to the choice of the density ξ, describe various classes of estimators improving upon δ_o.

Example 1. If we take $\eta(||x||^2) = 1$ in the definition of ξ, then (iv) is obviously satisfied, while (v) is satisfied for $p \geq 3$. In this particular case, we obtain Lemma 3.1 with $q = 2$.

Estimators which improve upon δ_o are given by:

$$\delta_2(X) = \frac{1}{\alpha K^p} \int_{||\tau|| \leq K} \delta_1(X - \tau)\, d\tau \quad , \; K \geq K_o \tag{3.41}$$

and we recognize again formula (3.33) (i.e., convolution with a uniform density).

Example 2. Consider $\eta(||x||^2) = 1/||x||^2$. By using polar coordinates in $\mathbb{R}^p$, we observe that condition (v) is satisfied for $p \geq 5$ (see also remark (1) below).

We can also prove that condition (iv) is satisfied:

$$\sup_{y \in \mathbb{R}^p} \int_{||x|| \leq r} \frac{1}{||x + y||^2}\, dx < \infty \; , \; (\forall)\; r > 0 \; . \tag{3.42}$$

To see this, perform a change of variables, by applying the $p \times p$ orthogonal transformation T, such that $T(y) = (||y||, 0, 0, \ldots, 0)$. We get

$$\int_{||x|| \leq r} \frac{1}{||x + y||^2}\, dx = \int_{||x|| \leq r} \frac{1}{(x_1 + ||y||)^2 + x_2^2 + \ldots + x_p^2}\, dx \; . \tag{3.43}$$

Then, transform $(x_2, x_3, \ldots, x_p)$ into spherical coordinates:

$$x_1 = u$$
$$x_2 = \rho \cos \phi_2$$
$$(3.44) \qquad x_3 = \rho \sin \phi_2 \operatorname{cox} \phi_3$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$x_p = \rho \sin \phi_2 \ldots \sin \phi_{p-1}$$

where $\phi_2, \ldots, \phi_{p-2} \in [0,\pi]$, $\phi_{p-1} \in [0,2\pi]$, and $\rho \in [0,r]$.

After this transformation, (3.43) becomes:

$$(3.45) \qquad Q \int_0^r \int_{-\sqrt{r^2-\rho^2}}^{\sqrt{r^2-\rho^2}} \frac{\rho^{p-2}}{(\|y\| + u)^2 + \rho^2} \, du\, d\rho$$

where Q is a constant.

After some simple calculations, this expression simplifies to:

$$(3.46) \qquad Q \int_0^r \rho^{p-3} \left[\tan^{-1} \frac{\|y\| + \sqrt{r^2 - \rho^2}}{\rho} - \tan^{-1} \frac{\|y\| - \sqrt{r^2 - \rho^2}}{\rho}\right] d\rho \ .$$

Since $-\pi/2 < \tan^{-1} a < \pi/2$, we observe that the integral in (3.46) is bounded by a constant depending only on r . This shows that condition (iv) is satisfied.

The class of estimators which improve upon δ_o is given by:

$$(3.47) \qquad \delta_2(X) = c \int_{\|\tau\| \le K} \frac{\delta_1(X - \tau)}{\|\tau\|^2} \, d\tau \ , \ K \ge K_o$$

where c is a constant of the order $1/K^{p-2}$.

Example 3. To generalize the previous example, we take

$\eta(||x||^2) = 1/||x||^s$ with $s > 0$ and we consider the corresponding truncated density ξ given by (3.34).

Without repeating the calculation which is similar to that in Example 2, we mention that classes of estimators better than δ_o are obtained in higher dimensions. More exactly, the critical dimension for inadmissibility depends on s : the estimators

$$(3.48) \qquad \delta_2(X) = c \int_{||\tau|| \le K} \frac{\delta_1(X - \tau)}{||\tau||^s} \, d\tau \ , \ K \ge K_o$$

have smaller risk than $\delta_o(X) = X$, for $p > s + 3$. Here c is a constant of the order $1/K^{p-s}$.

Remarks. (1) Observe that condition (v) of Theorem 3.2 can be stated in the equivalent form:

$$(3.49) \int_0^\infty t^{p-1} \eta(t^2) \, dt = \infty \quad \text{and} \quad \lim_{t\to\infty} t^{p-2} \eta(t^2) = \infty \ .$$

This can be proved by using polar coordinates.

(2) In the proof of Theorem 3.1, the estimator δ_1 which improves upon δ_o outside of a compact set, is explicitly given. By using this and Theorem 3.2, we can give a more explicit form for the estimator δ_2 , which is uniformly better than δ_o .

Consider X an observation from the density $f(x - \theta)$, and suppose that the hypotheses of Theorem 3.1 are satisfied. The loss function is given by (3.1), where we assume for simplicity that $c_i a_{ii} = Q$ = constant, for all $i = 1,2,\ldots,p$.

Then the estimator δ_1 becomes:

$$\delta_1(X) = \left(1 - \frac{\varepsilon}{Q\|X\|^2}\right) X \tag{3.50}$$

If we consider the "convoluting" distribution to be uniform in the ball with radius K, it can be easily shown that (3.36) becomes:

$$\delta_2(X) = X - \frac{\varepsilon}{Q\alpha K^p} \int_{\|y-X\|\le K} \frac{y}{\|y\|^2}\, dy\ . \tag{3.51}$$

This estimator is better than δ_o if $K \ge K_o$. The constant K_o depends (as it can be seen from the proof of Lemma 3.1) on various other constants, such as R_o, B, a, p. We can take

$$a = \varepsilon\left[2p - 4 - \frac{\varepsilon}{Q \min\limits_i a_{ii}}\right] \tag{3.52}$$

with $0 < \varepsilon < 2(p-2) \cdot Q \cdot \min a_{ii}$, and we take K, such that:

$$K^p - r^p + r^p (R_o - B)(2K + r)^2/a > 0\ . \tag{3.53}$$

In specific examples, the constants R_o, B can be calculated, or replaced by some appropriate bounds.

Formulae analogous to (3.51) can be obtained by using other truncated distributions, such as those given in the examples above.

3.3. Further developments

It is possible to extend the results of previous sections and to describe other classes of estimators which improve upon $\delta_o(X) = X$.

Consider X having a density of the location type $f(x - \theta)$ and consider the loss function (3.1) where, for simplicity, we assume that $c_i a_{ii} = c_i \operatorname{Var}_0(X_i) = Q$ (a constant), for all $i = 1,2,\ldots,p$.

We assume that the hypotheses of Theorem 3.1 are satisfied. Without loss of generality, we can take $c_i a_{ii} = 1$, $i = 1,2,\ldots,p$.

Theorem 3.3. *If we denote by*

$$\delta_1(X) = \left(1 - \frac{\varepsilon}{\|X\|^s}\right) X \tag{3.54}$$

with $s > 2$ *and* $\varepsilon > 0$, *then*

$$\delta_2(X) = \frac{1}{\alpha K^p} \int_{\|y\| \le K} \delta_1(X - y)\,dy \quad , \; K \ge K_0 \tag{3.55}$$

is a better estimator than $\delta_0(X) = X$, *in dimensions* $p > s$.

Proof: With the notations of Section 3.1 we have:

$$D(\theta) = 2\sum_i h_{ii}(\theta) - \sum_i c_i h_i^2(\theta) \ . \tag{3.56}$$

A simple calculation gives:

$$h_{ii}(\theta) = \frac{\varepsilon}{\|\theta\|^{2s}} \left[\|\theta\|^s - s\theta_i^2 \|\theta\|^{s-2}\right] \ . \tag{3.57}$$

Therefore, we obtain:

$$\begin{aligned} D(\theta) &= 2 \sum_i \frac{\varepsilon}{\|\theta\|^{2s}} [\|\theta\|^s - s\theta_i^2\|\theta\|^{s-2}] - \sum_i c_i \frac{\varepsilon^2\theta_i^2}{\|\theta\|^{2s}} \\ &= 2\varepsilon \frac{p}{\|\theta\|^s} - \frac{2\varepsilon s}{\|\theta\|^s} - \varepsilon^2 \sum_i \frac{c_i\theta_i^2}{\|\theta\|^{2s}} \quad . \end{aligned} \tag{3.58}$$

Since $c_i < A$ for all $i = 1,2,\ldots,p$, where A is a constant, we get:

$$D(\theta) \geq \frac{2\varepsilon(p - s)}{\|\theta\|^s} - \frac{\varepsilon^2 A}{\|\theta\|^{2s-2}} \quad . \tag{3.59}$$

Therefore:

$$\|\theta\|^s D(\theta) \geq 2\varepsilon(p - s) - \frac{\varepsilon^2 A}{\|\theta\|^{s-2}} \quad . \tag{3.60}$$

Since $s > 2$, it follows that:

$$\liminf_{\|\theta\| \to \infty} \|\theta\|^s D(\theta) \geq 2\varepsilon(p - s) > 0 \tag{3.61}$$

for $p > s$ and $\varepsilon > 0$, by applying Lemma 3.1 with $q = s$. This ends the proof.

<u>Note</u>. Observe that in the case $s > 2$ we need $\varepsilon > 0$, but ε is otherwise unrestricted.

If $s = 2$, by looking at the proof above, we see that we need $0 < \varepsilon < 2(p - 2)/A$.

If $2 < s < 3$ we get better estimators in dimensions $p \geq 3$;

otherwise the improvement upon $\delta_o(X) = X$ is obtained in higher dimensions.

From the results of Section 3.2 and 3.3 it is clear that estimators δ_2 which improve upon the best invariant estimator δ_o are obtained as the convolution of some estimator δ_1 which improves upon δ_o outside of a compact set, with a suitable probability density ξ in $\mathbb{R}^p$:

$$(3.62) \qquad \delta_2 = \delta_1 * \xi \quad .$$

An interesting problem would be to study and, if possible, to characterize the following class of densities in $\mathbb{R}^p$:

$$(3.63) \quad \mathcal{D} = \{\xi \,/\, R(\theta, \delta_1 * \xi) < R(\theta, \delta_o) \ , \ (\forall)\theta \in \mathbb{R}^p\} \quad .$$

Note that $\mathcal{D}$ contains suitable normal densities (see Brown (1975)), as well as spherically uniform densities, and truncated densities with properties (iv) and (v) in Theorem 3.2.

A characterization of $\mathcal{D}$ would be useful in order to find wider classes of estimators which improve upon the best invariant procedure $\delta_o(X) = X$.

An interesting problem, closely related to this, is whether any estimator which improves upon δ_o can be written as a convolution of some estimator which improves outside of a compact set, with a suitable p-dimensional probability density.

The answer at these problems, which is not known even when sampling from the multivariate normal distribution, would possibly give a better understanding of the structure of estimators which improve upon the classical, best invariant procedure.

BIBLIOGRAPHY

[1] BARANCHIK, A.J. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Stanford Univ., Technical Report No. 51.

[2] BERGER, J.O. (1975). Minimax estimation of location vectors for a wide class of densities. Ann. Statist. 3 1318-1328.

[3] BLACKWELL, D. and GIRSHICK (1954). Theory of Games and Statistical Decisions. John Wiley and Sons, New York.

[4] BRANDWEIN, A.C. (1979). Minimax estimation of location parameters for spherically symmetric distributions under general quadratic loss. J. Multivariate Anal. 9 579-588.

[5] BRANDWEIN, A.C. and STRAWDERMAN, W.E. (1978). Minimax estimation of location parameters for spherically symmetric unimodal distributions. Ann. Statist. 6 377-416.

[6] BRANDWEIN, A.C. and STRAWDERMAN, W.E. (1980). Minimax estimation of location parameters for spherically symmetric distributions with concave loss. Ann. Statist. 8 279-284.

[7] BROWN, L.D. (1975). On the admissibility of invariant estimators of one or more location parameters. Ann. Math. Statist. 37 1087-1136.

[8] BROWN, L.D. (1975). Estimation with incompletely specified loss functions (the case of several location parameters). J. Amer. Statist. Assoc. 70 417-427.

[9] BROWN, L.D. (1979). A heuristic method for determining admissibility of estimators - with applications. Ann. Statist. 7 960-994.

[10] FARRELL, R.H. (1964). Estimators of a location parameter in the absolutely continuous case. Ann. Math. Statist. 35 949-998.

[11] FERGUSON, T.S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

[12] GIRSHICK, M.A. and SAVAGE, L.J. (1951). Bayes and minimax estimators for quadratic loss functions. Proc. Second Berkeley Symp. Math. Statist. Probability 1 53-73. Univ. California Press.

[13] GHOSH, J.K. and SINGH, R. (1970). Estimation of the reciprocal of scale parameter of a gamma density. Ann. Inst. Statist. Math. 22 51-55.

[14] GHOSH, M. and MEEDEN, G. (1977). Admissibility of linear estimators in the one parameter exponential family. Ann. Statist. 5 772-778.

[15] GUPTA, M.K. (1966). On the admissibility of linear estimates for estimating the mean of distributions of the one parameter exponential family. Calcutta Statist. Assoc. Bull. 20 14-19.

[16] HODGES, J.L. and LEHMANN, E.L. (1951). Some applications of the Cramér-Rao inequality. Proc. Second Berkeley Symp. Math. Statist. Probability 1 13-22 Univ. California Press.

[17] JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Probability 1 367-379 Univ. California Press.

[18] KARLIN, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist. 29 406-436.

[19] KATZ, M.W. (1961). Admissible and minimax estimators of parameters in truncated spaces. Ann. Math. Statist. 32 136-142.

[20] LEHMANN, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

[21] PING, C. (1964). Minimax estimates of parameters of distributions belonging to the exponential family. Chinese Math. Acta 5 277-299.

[22] STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Probability 1 197-206 Univ. California Press.

[23] STRAWDERMAN, W.E. (1974). Minimax estimators of location parameters for certain spherically symmetric distributions. J. Multivariate Anal. 1-10.

[24] WALD, A. (1950). Statistical Decision Functions. Wiley, New York.

[25] WESLER, O. (1959). Invariance theory and a modified minimax principle. Ann. Math. Statist. 30 1-20.

[26] ZACKS, S. (1971). The Theory of Statistical Inference. Wiley, New York.

[27] ZIDEK, J.V. (1970). Sufficient conditions for the admissibility under squared error loss of formal Bayes estimators. Ann. Math. Statist. 41 446-456.

VITA

Name: Dan A. Ralescu

PII Redacted [redacted] [redacted]

Degrees:

- Licenta, University of Bucharest, Romania (1972)
- M.A. (Mathematics), Indiana University (1979)
- Ph.D. (Mathematics), Indiana University (1980)

Academic Employment:

- Researcher at the Central Institute for Management and Informatics, Bucharest, Romania (1972-1976).
- Visiting Professor at the Department of Mathematics, University of Paris, Orsay, France (Fall, 1976).
- Assistant Professor at the Department of Economics and Management, University of Dijon, France (1976-1977).
- Visiting Research Associate at the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley (Summer 1979).

Publications:

(1) "Fuzzy Sets and Their Applications", Ed. Technica, Bucharest, 1974 (in Romanian; with C.V. Negoita).

(2) "Applications of Fuzzy Sets to Systems Analysis", Wiley, New York, 1976 (with C.V. Negoita).

(3) "Introduction to Fuzzy Systems Theory", Ohm-Sha Co., Tokyo, 1978 (in Japanese; with K. Asai, K. Tanaka, T. Okuda, and C.V. Negoita).

(4) "Weak solutions and generalized solutions for the elasto-plastic torsion operator", Ann. Univ. Bucharest, 1975, pp. 17-28.

(5) "Comment on a comment on an algorithm that generates fuzzy prime implicants by Lee and Chang", Information and Control 30, 1976, pp. 199-201.

(6) "L-fuzzy sets and L-flou sets", Elektr. Informat. und Kybernetik 12, 1976, pp. 599-605.

(7) "Fuzzy subobjects in a category and the theory of C - sets", Journal of Fuzzy Sets and Systems 1, 1978, pp. 193-202.

(8) "Representation of fuzzy concepts; a survey of the applications", in M. Gupta, R. Ragade, and R. Yager (eds.): "Advances in Fuzzy Set Theory and Applications", North-Holland, Amsterdam, 1979.

(9) "Toward a general theory of fuzzy variables", Memo UCB/ERL, Electronics Research Laboratory, University of California, Berkeley, 1979.

(10) "The fuzzy integral", J. Math Analysis Applic. 75, 1980 (with G. Adams).